Examining Changes in Transit Passenger Travel Behavior through a Smart Card Activity Analysis

By

Carlos H. Mojica

Civil Engineer, Universidad de Los Andes (2004)

Submitted to the Department of Urban Studies and Planning and the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the degrees of

Master in City Planning
and
Master of Science in Transportation
at the

Massachusetts Institute of Technology
June 2008

Author ....
Department of Urban Studies and Planning
Department of Civil and Environmental Engineering
May 19th 2008

Certified by..
Mikel Murga
Research Associate, Department of Civil and Environmental Engineering
Thesis Supervisor

Accepted by..
Daniele Veneziano
Chairman, Department Committee for Graduate Students

Accepted by..
Langley Keyes
Chair, MCP Committee
Department of Urban Studies and Planning

Examining changes in transit passenger travel behavior through a Smart Card activity analysis

By

Carlos H. Mojica

Submitted to the Department of Urban Studies and Planning and the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the degrees of Master in City Planning and Master of Science in Transportation

Abstract

Transit passenger behavior is an area of major interest for public transportation agencies. The relationship between ridership and maintenance projects, however, is unexplored but increasingly relevant in the era of aging infrastructure. This thesis bridges this gap by analyzing changes in Smart Card activity for a sample of rail commuters during a large scale maintenance project in Chicago. Results show that between 8% and 11% of the passengers used the bus system as a commuting alternative while the majority of them continued using the train under deteriorated service conditions. Comparisons to a control zone show that between 2% and 7% of the commuters did not use transit for their trips. Using the observed results, we model the shift from rail to bus using a binary logit model. Implications of the findings are discussed.

Thesis Supervisor: Mikel Murga
Department of Civil and Environmental Engineering

Thesis Reader: P. Christopher Zegras.
Department of Urban Studies and Planning

ACKNOWLEDGMENTS

First, I want to thank my family for the unconditional life-long support I have received. I was very fortunate to be born in the heart of such a large and tightly knit group, where you feel supported, appreciated and respected for every decision you take. Special honors to my mom Carlota, my dad Carlos and my four sisters Angela, Marcela, Carlota and Diana. They are the reason I can be writing these words and I can only say thank you, thank you, thank you: You deserve all the credit too. My uncle Abel and my cousins Carlos Jose, Claudia, Ricky, Jimmy and Andres Felipe have also been an invaluable support while I have lived in the United States. Thank you too.

Second, I want to acknowledge the MIT faculty members that worked with me and had a particular impact in my learning process. Special recognition to Nigel Wilson and Mikel Murga for giving me the opportunity to join the Transit Research Group. Chris Zegras, for his willingness to work hard and for sharing his vision of transportation. Fred Salvucci, for his continuous display of passion for public service. Bish Sanyal, for his inmense support of planning ideas and for retro fitting the word bureaucracy. Carl Martland, for allowing me to be his Teaching Assistant. Ralph Gakenheimer, for his support during my first semester at MIT. Diane Davis for her insights on Mexico City. Other remarkable transportation faculty were German Lleras, at Universidad de Los Andes, who supported my application to MIT, Daniel Rodriguez, at UNC, who let me work with him in an exciting research about Bogota and Jose Gomez-Ibanez at Harvard for his interest in Transmilenio. Special thanks to Ginny Siggia for helping me deal with the bureaucracy at MIT.

Third, the staff members at the Chicago Transportation Authority who made this research possible. I had an outstanding internship experience and loved the city in the summer thanks to your support. Special recognition to Alex Cui for providing the datasets for this research. Also, Aimee Lee, Jeffrey Busby, Jessica Vargas and Jeffrey Sriver for heading the Strategic Planning Division; Jennifer Wieland and Jackie Kapusnik for her fun stories about the CTA. Best wishes to President Huberman and his staff in achieving his goals for the future of the agency.

Fourth, my friends. I have made so many good friends in these past three years that I can not acknowledge all of them properly. My buddies at 1-235 and the TRG, who have been a constant support during thesis writing and will be there for a long time: David Uniman, Hazem Zureikat, Harvey Scorcia, Marty Milkovitz, Joshua Nelson, Sean Sweat, Mike Hanowsky, Catherine Seaborne, Candace Brakewood, Clara Yuan, Michael Frumin, Michael Kay, Joanne Chan, Tegin Teich, Brittany Montgomery, André Carrel, Tanachai Limpaitoon, Zhan Guo, Jinhua Zhao, Shan Jiang and DCBS. My good friends at DUSP, who are a source of fine stories in classes and in the bars: Rodrigo Diaz, Luis Flores, Miguel Paredes, Eric Brewer-Garcia, Mariana Arcaya, Kara Reeve, Sophie Martin, Francesca Napolitan, Sabra Loews, Zeina Saab, Marilia Magalhaes, Marisa Arpels, Patricia Acosta, Juan Rodriguez, Peter Banks, Jeff Schwarz and Luis Montoya. The Colombian Students Association, for making MIT a nicer and warmer place: Francisco Ruiz, Fabio Gordillo, Arturo Dominguez, Angelica Osorno, Juan Manuel Pedraza and Antoine Dumit. My friends in Washington who were just great during my internship at the IADB: Nicolas Estupinan, Felipe Targa, Martin Kunik, Bruno Laranjeira, Agustina Paglayán, Mauro Galeotti and Fernanda Almeida. And my friends in Colombia who have been there for a very long time: Francisco Tobar, Ricardo Baez, Giancarlo Santos, Andres Zapata, Luis Lopera, Daniel Beltran, Jose Pinzon, Valentina Duque, Felipe Castro and Ana Carrillo.

Finally, a special acknowledgement to Diana Briceño for her love and patience during the last three years. I am very excited to be right here right now and I am looking forward to spend a long time together. Thanks to Diana, Eduardo, Maria Antonieta, Silvia, Luis and Diego who have embraced me in their family too.

Carlos H. Mojica
May 21st, 2008

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

This thesis is part of the MIT-CTA research collaborative efforts and has two chief motivations:

First, the general goal of making the CTA a more competitive agency by incorporating new technologies into their internal planning processes. In particular in the objective to incorporate the revealed information from the system-wide use of the Smart Cards into the transportation planning context.

Second, a more specific goal is to answer the question of how do the maintenance and expansion projects affect the travel decisions of the agency's customers. Maintenance is becoming increasingly important as the system ages and in the upcoming years will likely draw even more attention.

This thesis reconciles both motivations by examining changes in traveler's behavior during a large scale infrastructure maintenance and expansion project in Chicago by analyzing the records of individual Smart Cards along different periods of time.

This first chapter summarizes the objectives, approach and findings of this thesis.

1.1 Research objectives

a. Assess the ridership implications of the Brown Line Capacity Expansion Project (BLCEP). To achieve this objective the following questions are posed. What kind of data sources are available for the Chicago Transit Authority's staff to make a ridership assessment? What are the advantages and disadvantages of each? How can the limitations of aggregate level counts be overcome to produce an estimate of the project-related ridership impacts and, what was the extent of these impacts on the different stages of the project?

b. Develop a methodology to collect, process and select records from available automated sources -Automated Fare Collection (AFC) and Automatic Vehicle Location (AVL) systems- to produce a longitudinal 'panel' data based on Smart Card activity. To achieve this objective the following questions are posed: What is the extent of the use of Smart Cards as fare media in Chicago? Is there existing research on the potential applications of these devices for transit planning purposes? How can the CTA keep track of user-specific activity over different periods of time and capture behavioral changes? What are the drawbacks of using this approach?

c. Understand how passengers change their travel habits when service decreases due to announced maintenance projects and develop a framework to forecast the impact of maintenance projects on passenger ridership. To achieve this objective the following questions are posed: What are the observed changes in travel behavior when comparing two time periods for a panel of individuals? What were the trip characteristics of each individual? How good were his/her other alternatives? What analytical framework is suitable to study and model these behavioral changes? How can the CTA take advantage of these findings for future planning efforts?

1.2 Thesis structure and approach

The chapters of this thesis are organized in the following way:

Chapter two presents a literature review that explores the concept of Smart Cards and its applications to the transit industry. The chapter shows that the use of Smart Cards in transit agencies is widespread under different fare policies and contractual schemes. However, in the case of the CTA there is still room for improvement in terms of fare media market share and spatial coverage. The chapter introduces previous research that has been done in this regard and builds from findings and methodologies developed by the CTA and other fellow MIT researchers.

Chapter three presents a descriptive and quantitative analysis of the Brown Line Capacity Expansion Project (BLCEP) from the point of view of the transit agency and the changes in rail system ridership. It starts by introducing the project in the context of the needs for infrastructure maintenance and expansion and describes how the CTA planned and announced the contingency plan to counter the inconveniences created by such project. This chapter uses quantitative information from the automated passenger count system to analyze the changes in rail ridership levels. Preliminary evidence described in this thesis shows a change in modal preferences during the different stages of the Brown line capacity expansion project, likely attributable to a decrease in the train level of service.

Chapter four develops a methodology to use the Smart Card records to identify travel patterns and create user specific origin destination profiles. This methodology merges data from the Automated Fare Collection (AFC) and Automatic Vehicle Location (AVL) systems with geographical data for the Chicago transit network and the location of smart card users. A step by step procedure is described and implemented to study the travel patterns of rail commuters who live around selected stations affected by the BLCEP.

Chapter five attempts to quantify changes in travel behavior by examining individual smart card activity. Two cross sections are examined: before and during the BLCEP. Several variables are examined and evidence is shown of changes in transit modal selection for segments of the studied population.

Chapter six explains the observed changes under the random utility framework. A binary logit model is developed to forecast modal shifts from rail to bus. The model produces reasonable estimates for coefficients and a sample forecast is made to illustrate its use for bus service and budget allocation.

Chapter seven summarizes the findings of this thesis and presents suggestions for its application and future research projects.

1.3 Contribution to the Chicago Transit Authority

This thesis provides three specific contributions:

1. A methodology to account for changes in rail boardings solely based on station taking into consideration seasonal changes. The use of the methodology is illustrated in the case of the recent BLCEP
2. A step by step methodology to merge the Smart Card registration data with AFC and AVL records to produce information about individual travel behavior. This methodology is applied in two different time periods to customers affected by the BLCEP.
3. A discrete choice model that represents modal shift of rail passengers subject to a decrease of train level of service. The model is applicable to predict changes in ridership at the station level by using data from residents' travel activity.

2 Smart Cards and transit applications

The second chapter of this thesis is devoted to explaining the concept of a Smart Card, its characteristics and its use within the transit industry. Special consideration will be given to the Chicago Transit Authority's use of the Chicago Card and Chicago Card Plus programs. The understanding of these concepts and background should provide the reader a general perspective on the type of technology that is being be discussed and the type of application that this research is developing.

2.1 Description

The Transit Cooperation Research Program (2003)[1] defines the Smart Cards as plastic pocket-sized cards with an embedded integrated circuit. Two types of cards can be identified, based on their circuit technology. A first group has only one memory chip, which can store pre-programmed data which can not be modified. A second group of cards has a memory chip and a microprocessor which allows to reprogram the card multiple times. This latter group is likely to be used for transit and other potential applications, where users constantly require to add value to cards as they use them

Smart Cards are designed with contact or contactless interfaces by numerous industries. On one hand, industries such as the cellular telephone industry used contact smart cards for their GSM technology. As its name implies, these cards require physical contact between a reader and the chip of the card. On the other hand, contactless cards have had an impact on industries since their development in the 1990s and have been widely used for security purposes, retail payments, banking transactions and recently in the transit industry as well.

[1] TRCP 94 (2003) Fare Policies, Structures and Technologies: Update

2.2 Use in transit agencies

The use of Smart Cards in the Transit industry has expanded in the last twenty years. The first tests were conducted in European agencies, and in the Asian continent. Nowadays, numerous transit systems worldwide include Smart Cards as one of their fare media. Providing, distributing and maintaining the corresponding infrastructure is commonly contracted to an outside vendor. This vendor usually serves the function of integrating the agency's fare policies with the customer's trip experience. Table 1 shows a list of U.S. agencies that by 2003 were using, or planning to introduce, Smart Cards as their fare media payment option

City/Agency	Type of service	Integrator / vendor	Status
Los Angeles/LACMTA (UFS)	Regional farecard	Cubic	Contract awarded rollout planned in 2004
San Diego/MTDB	Regional farecard	Cubic	Contract awarded rollout planned in 2005
San Francisco/MTC(TransLink)	Regional farecard	ERG	Pilot completed mid-2002; additional cards/equipment ordered mid-2003
Ventura County/VCTC	Regional farecard	ERG	Implemented 2002
Washington-Maryland-Virginia/WMATA (SmarTrip)	Regional farecard	Cubic/GFI	In use on MetroRail contract awarded for rest of region
Delaware/DelDOT	Regional farecard	NA*	Under development
Miami-Ft. Lauderdale-Palm Beach/MDTA-Tri Rail (UAFC)	Regional farecard	Cubic	Contract awarded 2002
Orlando/Lynx (ORANGES)	Multimodal integration	TTI	Under development
Atlanta/MARTA	Regional farecard	Cubic	Contract awarded 2003 card rollout in next phase
Chicago/CTA	AFC option (also regional)	Cubic	Completed
Boston/MBTA	AFC option	Scheidt & Bachmann	Contract awarded 2003
Las Vegas/Monorail	New fare system(new service)	ERG	Contract awarded 2002; transit service to open 2004
Minneapolis-St. Paul/Metro	Transit New fare system	Cubic	Contract awarded; rollout planned mid-2003
Newark/PANYNJ & NJT(SmartLink)	AFC option	Ascom/ASK	Pilot implemented 2001
New York City-NJ/PATH	AFC option	Cubic	Contract awarded 2002

Philadelphia/PATCO	New fare system	NA	Under development
Houston/METRO	AFC upgrade	Cubic	Contract awarded 2002
Seattle-Puget Sound/KC Metro	Regional farecard	ERG	Contract awarded; rollout

Table 1: Current and planned U.S. transit smart card programs
Source: TCRP 94 Fare Policies, Structures and Technologies: Update (2003)

The justification for upgrading the fare payment media to Smart Cards technology is based on the benefits that are transferred to the customers and to the agency as well, as identified by Yi (2006)[2]:

Potential benefits for customers in terms of:

a. Ease of usage, as the technology is friendly to customers

b. Faster boarding times by reducing the transaction times

c. No need for exact change, as the system charges the exact value from the card

d. Enhanced security and durability, when compared to magnetic tickets.

Potential benefits for agencies in terms of:

a. Lower dwell times -on buses with readers- improving service reliability

b. Reduction in cash handling and improving the security of transit revenues

c. Reduction in fare evasion and fraud, when compared to other fare media

d. Better performance, compared to magnetic tickets, in terms of failed transactions.

e. Information about passenger behavior that can be used for planning purposes

Yi also presents a set of additional risks and drawbacks to the transit agencies by using Smart Cards.

a. The potential loss of private data and customer personal information

b. The monetary cost of each single card and the associated cost for the customer.

c. The risks associated with using a rising technology in a traditionally lagging industry

2 Yi, Hong (2006) transition to Smart Card technology: How transit operators encourage the take-up of Smart Card technology.

2.3 Use in the Chicago Transit Authority

The Chicago Transit Authority (CTA) operates the second largest public transportation system in the United States. In addition to the city of Chicago, more than 40 suburbs are served as well by the CTA. More than 1.6 million rides are taken on an average weekday on a system that combines buses and trains. The fare media options that customers have are cash (only on buses), magnetic cards, magnetic passes, and Smart Cards. Table 2 summarizes the CTA's fare structure.

Fare media	**Full Fares**
Cash	$ 2.00
Full Fare Transit Card - Bus	$ 1.75
Full Fare Transit Card – Rail	$ 2.00
Full Fare Chicago Card/Chicago Card Plus	$ 1.75
Chicago Card/Chicago Card Plus plus Bonus	10% for $20 added
Transit Card, Chicago Card and Chicago Card Plus Transfer (bus and rail)	$ 0.25
1-day Pass	$ 5.00
2-day Visitor Pass	$ 9.00
3-day Visitor Pass	$ 12.00
5-day Visitor Pass	$ 18.00
Full Fare 7-day Pass	$ 20.00
Full Fare 30-day Pass	$ 75.00
Fare media	**Reduced Fares**
Cash	$ 1.00
Transit Card (bus and rail)	$ 0.85
Transit Card Transfer (bus and rail)	$ 0.15
Reduced Fare 30-day Pass	$ 35.00

Table 2: 2006 Fare Structure
Source: www.transitchicago.com

The CTA implemented a Smart Card program which features two different types of cards. The Chicago Card (CC) and the Chicago Card Plus (CC+)..

The Chicago Card is a regular smart card which has the ability to store a value and be read by all the system readers, turnstiles and fareboxes. The type of microchip in the card allows the customer to add value to it in any of the CTA vending machines or in any point of sale. The customer can register some of his personal information in order to protect the stored value on the card in case it is lost or destroyed.

The Chicago Card Plus has the same physical characteristics (although under a different external appearance) as the Chicago Card, but is managed through an online account. This permits linking the card with a bank account so as to maintain a given balance online. It also allows the option to transfer funds from a bank account to reloading the card automatically when its balance runs out. The Chicago Card Plus can also be activated as of a monthly pass, providing unlimited rides for 30 days . The balance of the Chicago Card Plus is stored in a virtual account to prevent data losses if the card is lost or stolen. Table 3 summarizes the main differences between the CC and CC+ programs.

Feature	Chicago Card (CC)	Chicago Card Plus (CC+)
Adding Value	Only cash at CTA Vending Machines or off-site point-of-sale devices	Only credit cards or Transit Benefits dollars at www.chicago-card.com
Checking Value	Check at CTA vending machines or off-site point-of-sale devices	Check online or by phone calling CTA Customer Service
Fare Types	One choice: Pay-per-use	Two choices: Pay-per-use or 30-day pass
Registration	Optional	Required including e-mail address

Table 3: Key differences between Chicago Card and Chicago Card Plus.
Adapted from: Vargas Astaiza and Minser (2006)[3]

In terms of market share across fare media, the Smart Cards are not the dominating option in Chicago. By June of 2007 the CC and CC+ cards only accounted for 16.5% of all the system rides (CTA, 2007)[4]. Examination by mode reveals that this share corresponds to 27.6% in the rail system and 10.9% for bus rides. Table 4 summarizes fare media usage in the CTA:

[3] Vargas Astaiza, J. and Minser J. (2006) Expanding CTA's Smart Card Program: Market Research Efforts
[4] Chicago Transit Authority, Fare Media Report, June 2007

Fare media technology	Bus System	Rail System	CTA Total
Smart Card	10.9%	27.6%	16.5%
Magnetic transit card	67.5%	72.2%	69.1%
Cash	9.1%	0.0%	6.0%
Manual Counts*	12.6%	0.1%	8.4%

Table 4: CTA fare media market shares
Source: Fare media report June 2007
* unidentified fare media

The relatively low share of Smart Cards in the CTA can be explained by diverse reasons, such as:

a- The inability of the cards to be used as a 7-day passes (both CC and CC+) and as a 30-day pass (CC). This is due to both contractual and a technological constraints, which the CTA acknowledges and plans to overcome.

b- The initial limitation of reloading points in the rail stations. When the program was started, the rail stations (and other few locations in the city), constituted the only points to reload the Smart Cards, leaving the bus customers under in a comparative disadvantage.

c- The higher coverage if the rail system in denser and wealthier areas of the city. The Smart Card technology is more likely to be adopted by a population group that has a better access to credit cards and internet service, two key components of the CC+ program.

d- Lack of information among the base of CTA customers about the benefits of using Smart Cards

According to Vargas Astaiza and Minser (2006)[5], the CTA has already implemented three major policy decisions to foster the use of Smart Cards in the system. First, the Go Lane Program, where an express boarding lane is designed in the entrance of buses to facilitate the flow of Smart Card users. In addition to the fare box, these buses have an independent card reader for Smart Cards. Similarly, some rail stations have also dedicated one exclusive turnstile for CC and CC+ users. The objective is to attract new customers towards Smart Card by providing them faster access times to the system in a

[5] Vargas Astaiza, J. and Minser J. (2006) Expanding CTA's Smart Card Program: Market Research Efforts

similar way as toll highways provide express lanes to vehicles with toll transponders. The pilot program started in June of 2005. By 2007, 51% of the bus fleet was equipped with independent card readers and 8 rail stations with exclusive turnstiles. A second initiative, called the Touch-n-go Program started on December of 2005. This initiative aimed to expand the points where Smart Cards can be reloaded. Originally, reloading was limited to rail stations, but under this policy, reloading points were created in retail stores and currency exchange stores. Over 65 different points in the City of Chicago were opened for customers to buy or reload Smart Cards by paying the value at the cashiers. The program has expanded and currently more than 200 points are open for customers in different neighborhoods. A third initiative was related to changes in Fare Policy: by January 1st of 2006, the CTA had a system-wide fare increase, as shown in Table 5; however, the fare for boardings made with CC or CC+ did not increase. This policy also aimed to discourage the use of cash in the system by eliminating the cash transfers.

Fare media Technology	Before Jan. 1, 2006		After Jan. 1, 2006			
	Both Rail and Bus		At Rail Station		On Buses	
	Full Fare	Transfers	Full Fare	Transfers	Full Fare	Transfers
Chicago Cards	$1.75	$0.25	$1.75	$0.25	$1.75	$0.25
Transit Cards	$1.75	$0.25	$2.00	$0.25	$1.75	$0.25
Cash	$1.75	$0.25	$2.00*	Not issued	$2.00*	Not issued

Table 5: Changes in fare media policy
Source: www.transitchicago.com

The CTA also provided additional incentives by eliminating the initial $5 cost of the CC and CC+ between December '05 and January '06. This incentive, added to the increase in fares for non-Smart Card users, spiked the market share for the Smart Cards to 28% in June of 2006. Since then, the share has oscillated between 24% and 27% of total boardings as seen in Figure 1.

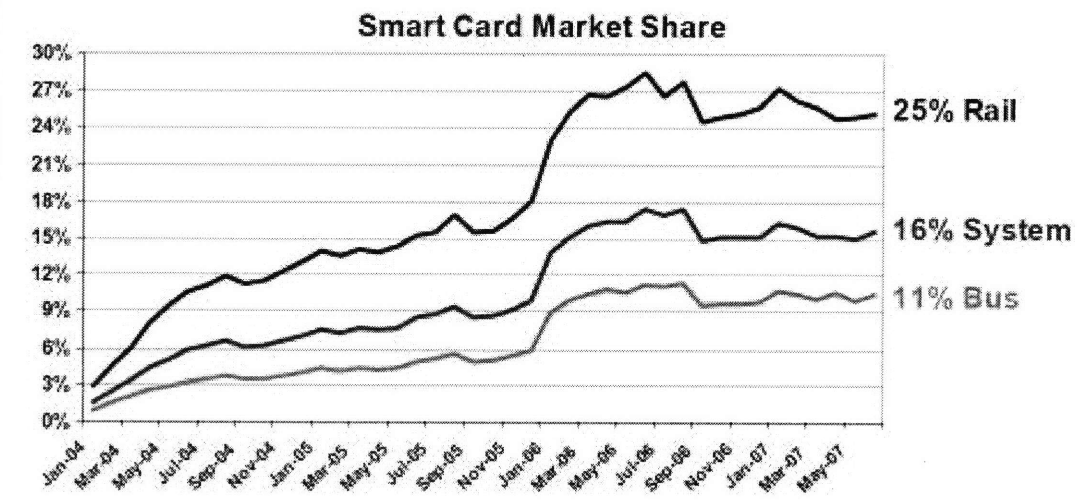


Figure 1: CTA Smart Card market share 2004-2007
Source: CTA Smart Card market: What do we know?[6]

However, this represents still a low market penetration` for Smart Cards compared to other transit agencies in the world. As seen in Table 6, the use of Smart Cards is more intense on agencies that have given other uses to the Cards and offer more stringent disincentives to the use of other fare media. Nevertheless, current policies on the CTA aim to increase usage in the medium and long term.

City/Transit agency	Uses	Smart Card Market share
Hong Kong (2006)	Transit, parking, retail, identification	MTR:90%
Washington DC Metro (2006)	Transit, parking	Rail: 60% Bus 18%
Singapore (2006)	Transit, retail, identification	Rail: 96% Bus: 90%
Chicago (2007)	Transit	Rail: 25% Bus: 11%
London (2006)	Transit	Rail: 68% Bus: 50%

Table 6: Smart Card market share in selected cities
Source: Adapted from Yi (2006)[7]

[6] CTA, Market research.(2007) CTA Smart Card market: What do we know?.
[7] Yi, Hong (2006) transition to Smart Card technology: How transit operators encourage the take-up of Smart Card technology.

Smart Cards promise to be an important source of information for planning purposes within a transit agency. Customers are required (in the case of CC+), or can opt (in the case of CC), to provide personal information when they acquire their card. This opens a window for the agency to use the data contained in the Smart Cards and link it to the Automated Fare Collection (AFC) system for transit planning, market research and operational purposes, as seen in Table 7.

Market Research and Service Planning
Analysis of demographics (e.g. age, gender, income) of riders by route/station
Analysis of travel characteristics (e.g. frequency of use, transfer patterns, etc.) of riders by route/station
Analysis of travel by riders with particular demographic characteristics
Analysis of travel by riders with particular travel patterns
Analysis of demographics of riders making particular trip patterns
Analysis of travel characteristics of riders making particular trip patterns
Analysis of the spatial coverage of CTA's system
Analysis of changes in travel patterns over time by people with particular demographics
Analysis of changes in travel patterns over time by people with particular travel characteristics
Analysis of the demographics of riders by time of day
Analysis of the transportation characteristics of riders by time of day
Identification of individuals for detailed survey or focus groups
Analysis of demographics of riders using particular CTA services (e.g. express, limited stop, night owl)
Development of a mailing list for public meeting notices
Travel Demand Forecasting
Provision of a large sample transit "travel diary", including demographic data
Study of travel changes as reactions to fare changes (elasticity) by demographics
Analysis of changes in travel behavior as a result of changes in level of service
Operations
Development of a mailing list for service change announcements
Development of an e-mail list for delays and emergency detours
Pricing/Fare Policy
Analysis of complete trip-making patterns to evaluate new fare products
Evaluation of the feasibility of a trip frequency-based discount or a "guaranteed best fare" policy
Analysis of travel by fare category
Study of price-elasticities by demographic characteristics
Marketing
Identification of distinct market segments among CTA riders
Targeting of marketing information to the most appropriate users
Identification of market segments with low penetration
Use of demographic database to conduct targeted surveys
Advertising
Development of route/station demographic profiles to identify target locations for particular CTA advertisers and set

advertising rates accordingly
Table 7: Potential applications of Smart Card database linked to AFC system Adapted from Transystems Corporation (2003) [8]

Applications of Smart Cards will become more useful as the agency can collect more detailed information about the particular characteristics of the customers. In particular, additional information regarding demographics, socio-economic conditions and other non-transit travel patterns will enrich the analyses made within the transit agency. Currently, there is no mechanism to link Smart Card activity information with other personal information except the reported address, but a incentive-based mechanism could be developed in a future for planning purposes.

2.4 Recent research on CTA Smart Cards:

Despite limitations in sample sizes and detailed passenger information, a fair amount of research has been dedicated to the study of Smart Cards in the CTA for market research and planning purposes. This thesis builds upon prior research to explore new applications from this rich database.

The CTA has completed several research studies to understand the perception that customers have on the CC and CC+. By means of analytical studies the CTA (2007)[9] found that CTA smartcard penetration and market share has steadily increased since its implementation in 2003, and most notably after the January 2006 incentives on fare policy, mentioned above. This study also used focus groups and conducted surveys, finding that CTA smartcards can attract a different customer market than the other fare media options. For instance, in comparison with overall CTA customers, CC and CC+ customers are more likely to be between the ages of 25 and 34, Caucasian, more affluent (household incomes above $60.000), more educated and employed full time. In terms of location, Smart Card customers are more likely to live in the North of Chicago. Finally, this study also found current CTA Smart Card users to be highly satisfied with their

[8] Transystems Corporation (2003) Memorandum: CTA SmartCard Traveler Database Feasibility
[9] CTA, Market research.(2007) CTA Smart Card market: What do we know?.

cards: Smart Cards are appreciated chiefly because of faster boarding times and lower fares. Non-Smart Card users reported not knowing enough about smartcards, the $5 initial cost and the lack of passes and reduced fare options as a barrier to switching to Smart Cards. Following these intensive market research efforts, the CTA is developing policies to increase the penetration of the CC and CC+ fare media such as those previously mentioned in section 2.3.

On the other hand, MIT has completed several research reports in conjunction with the CTA that are closely related to smart cards. Two main research trends can be identified. A first trend attempts to use the raw data from the AFC and the AVL systems for planning purposes; For example, in Zhao (2004)[10] and Wilson, Zhao and Rahbee (2005)[11] integrated these two data sources to infer the rail origin-destination (OD) matrix from the origin-only rail trip data. Furthermore, he studied rail path choice by employing discrete choice models to examine revealed public transit riders' travel behavior based on the inferred OD matrix and transit network attributes. In further work, Cui (2006)[12] developed a similar OD framework for the CTA bus system.

A second research trend is related to the explanation and use of the customer-associated data that can be inferred from the use of Smart Cards. Utsunomiya, Attanucci and Wilson (2006)[13] developed an analysis of the walk accessibility and usage patterns of Smart Card holders at the Chicago Transit Authority during September 2004. Their analyses included walk access distance, frequency and consistency of daily travel patterns, and variability of smart card customer behavior by residential area. Gupta (2006)[14] continued this work by studying other travel parameters such as frequency, time consistency, access distances, and route variability. Data from two weeks in September

[10] Zhao, Jinhua (2004) The Planning and Analysis Implications of Automated Data Collection Systems: Rail Transit OD Matrix Inference and Path Choice Modeling Example. MIT
[11] Wilson, Zhao and Rahbee (2005) The potential impact of automated data collection systems on urban public transport planning.
[12] Cui, Alex (2006) MIT Master's Thesis
[13] Utsunomiya, Attanucci and Wilson (2006) Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning
[14] Gupta, Saumya (2006) Understanding Transit Travel Behavior: Value added by Smart Cards. MIT thesis

2004 and September 2005 were used, at a time when the usage levels of the Smart Cards in Chicago were still very low.

This thesis continues those trends and focuses on of using data from the Smart Cards for specific project evaluation purposes. Specifically, this research explores how to use this rich data set in the service planning and ridership forecast stages of infrastructure maintenance projects.

3 Rail maintenance and expansion projects in Chicago

This chapter presents the context of the maintenance and expansion projects in the Chicago Transit Authority, with particular emphasis on the Brown Line Capacity Expansion Project (BLCEP). The project's characteristics and service implications are described and it presents the way the CTA dealt with them in terms of customer outreach and changes in existing service. The final section presents an estimation of the global rail ridership impacts of the project by examining the number of boardings in the Brown Line at different time periods.

3.1 Maintenance and expansion of the CTA rail system

The rail infrastructure of the CTA is an aging system with more than one hundred years of operation. By 2007, it featured 8 different lines covering 106 miles and 144 stations. It provides a direct connection for most areas of Chicago with the Central Business District (CBD), where most of the lines meet in the Loop. By June of 2007, average weekday ridership was estimated in 530,693 passengers[15] and station boardings recorded 161,966,231 million passengers in the year 2006[16]. The Appendix shows a full map of the CTA rail system.

However, the system is currently under-funded to provide the proper maintenance and bring it to a "state of good repair". Estimates of the Regional Transportation Authority show capital needs of 6.3 billions in the next 5 years to maintain bus and rail rolling stocks, track and subway structures, signaling and communication systems and other facilities[17]. Additional investments to enhance (US$ 328.9 million) and expand (US$ 655 million) the system are being planned to improve the level of service and its safety. Under current funding schemes, capital investments are subjected to cost-reduction

[15] CTA Rail ridership report by line and branch. June 2007
[16] CTA Rail ridership report by line and branch. December 2006
[17] RTA, Moving beyond congestion (2007)

exercises due to large operational deficits. By 2007, the current fare collection and other local revenues accounted for only (US$ 541 million) of the total operational expenditures (1.08 billion) while the remaining expenditures were partially covered by funding of the Regional Transportation Authority (RTA) (480 million). The expenses not covered by the two mentioned sources often rely on transfers of capital funds.[18]

This dire picture of large capital needs, under a tightly constrained budget, shows how important maintenance and expansion projects will become in the future. The CTA experienced one major incident in July 2006, suspected to be caused by poor infrastructure maintenance, when a Blue Line train derailed and sparked a fire in the Dearborn subway[19]. Although it did not cause any fatalities, victims filed lawsuit, with one of them recently settling for US$ 1.25million in April of 2008[20]. This incident triggered immediate action to improve maintenance efforts in the Blue Line to bring it to a state of good repair. Recently, the press[21] echoed an incident in which passengers decided on their own, and against the advice provided by the public address system, to abandon a stranded Blue Line train in the Dearborn tunnel. This risky and unplanned evacuation underscores the current challenges faced by the CTA.

For transportation planning purposes, it is relevant to identify the continuum of engineering approaches to rail infrastructure maintenance: This continuum exists because repair works bring inconveniences to customers, and the agency has to establish a trade-off between improving the efficiency of the project, by sacrificing customer's satisfaction, and extending the project longer and its cost, in order to smooth out the impact on users.

The CTA is currently undergoing Blue Line repairs by closing portions of the line on the weekends to allow construction crews to repair the tracks. This approach represents a

[18] CTA, 2007 Operating budget
[19] Chicago Tribune, July 12th 2006
[20] Chicago Tribune, April 15th 2008
[21] Chicago Sun-times, April 15th, 2008

customer impact minimization, where the majority of the trips are still allowed on the line while supplementary services -such as shuttle buses- are provided to keep connectivity. On the other hand, the Green Line rehabilitation project closed the line completely in 1994 for a period of two years in order to repair the elevated tracks. The decision was highly controversial as it was announced to the public only one month before the closure and included a reduction in the number of stations once reopened.[22] After being reopened in 1996, ridership levels did not climbed back above pre-closure levels until year 2000.[23]

3.2 BLCEP Project description

The Brown Line Capacity Expansion Project (BLCEP) is a $530 million dollar project envisioned to provide rail stations with accessibility for the disabled and increased capacity of the line by enabling the use of eight car trains. Previously, the system was running on a six car operation, mainly restricted by the size of the platforms. The project also includes improvements in the stations' interior design to reduce crowding by widening staircases and adding more turnstiles.

The infrastructure of the Brown Line is over one hundred years old and serves a portion of Chicago's northwest communities. It is a line that serves more than 66.000 customers each weekday. It has experienced significant ridership increases in the last decades, showing the highest growth rates across rail lines in the CTA[24]. The BLCEP corresponds to a planned infrastructure response to the increased demand in the line.

[22] Chicago L.org. http://www.chicago-l.org/history/CTA4.html
[23] CTA website, Press release 10/17/01
[24] www.ctabrownline.com

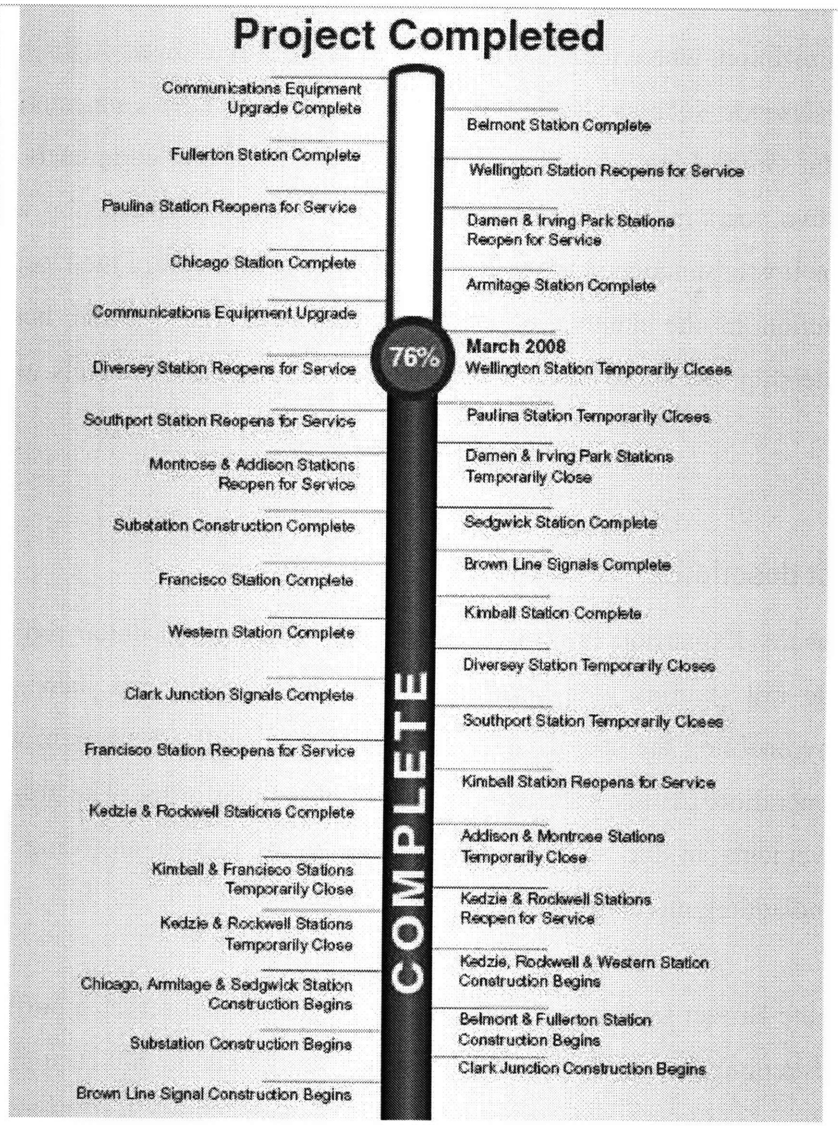


Figure 2: Calendar of activities of BCLEP.
Extracted from www.ctabrowline.com 04/25/2008

The BLCEP includes work on 18 Brown Line stations, signal improvements and enhancements to the energy substations. Work started in 2006 and is expected to finish by December 2009. As it can be seen in Figure 2, the project is planned as a 'staged' engineering project. This means that the station closures are strategically programmed to be closed to the public in stages, as opposed to closing all the line at the same time. By April 2008, the project was 76% complete.

Due to the expected negative impact that the project would have on line-surrounding residents, the CTA held numerous community meetings and made an extensive outreach effort to inform the affected residents. These efforts included holding public discussions before advertising bid packages, publishing extensive information -in the form of customer alerts and flyers- of the construction schedule and developing an outreach plan to aid impacted local businesses.[25]

Funding for the project was made available through the 2006 Federal Transportation Appropriation and the State of Illinois' Illinois FIRST program. The federal contribution was US$ 423.1 million[26] while the State funds provided additional US$ 102.6 million.[27]

3.3 Changes in operation

The construction work of the BLCEP induced changes in rail and bus operations to allow service in the Brown Line while the project is underway. For particular station closures, changes were limited to the temporal elimination of stops in those stations during a span of 6-12 months, depending on the particular construction schedule. These stations can be seen in Figure 3.

However, major changes were introduced in April 2nd of 2007 when the Brown, Red and Purple lines entered the so-called "Three track operation". The expansion of the Belmont and Fullerton stations required moving one of the four tracks that are shared by these three train lines in order to complete the upgrades, as it can be seen in Figure 4. The practical impact of this restriction is a reduction in throughput, which in turn means longer trip times and fewer scheduled trains.

[25] CTA Project Presentation 03/01/07. Olson Auditorium.
[26] Chicago Transit Authority website press release 4/13/04.
http://www.transitchicago.com/news/archpress.wu?action=displayarticledetail&articleid=119462
[27] State of Illinois website,
http://www.illinois.gov/PressReleases/ShowPressRelease.cfm?RecNum=1076&SubjectID=25

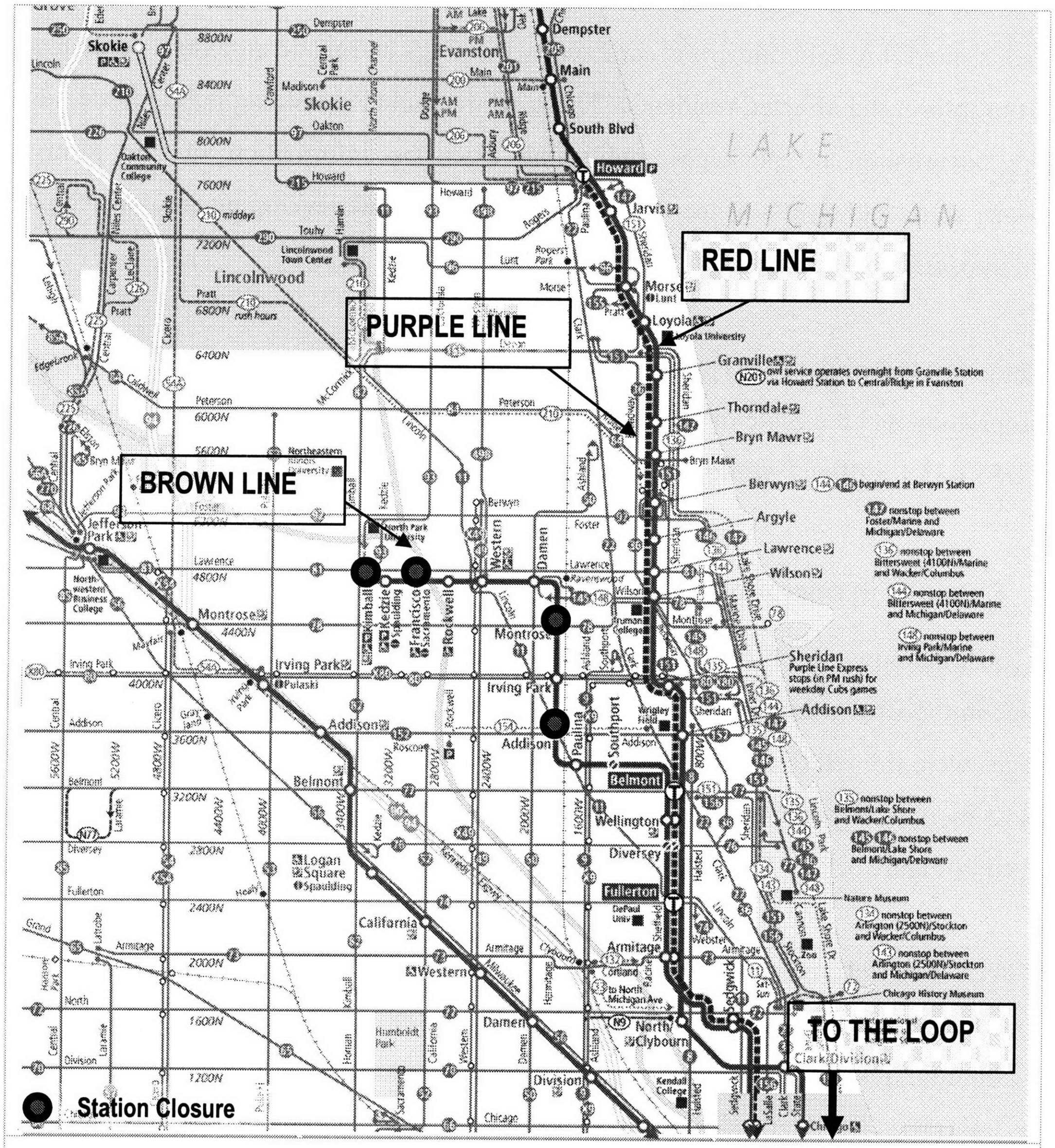


Figure 3: CTA North System Map
Source: www.transitchicago.com

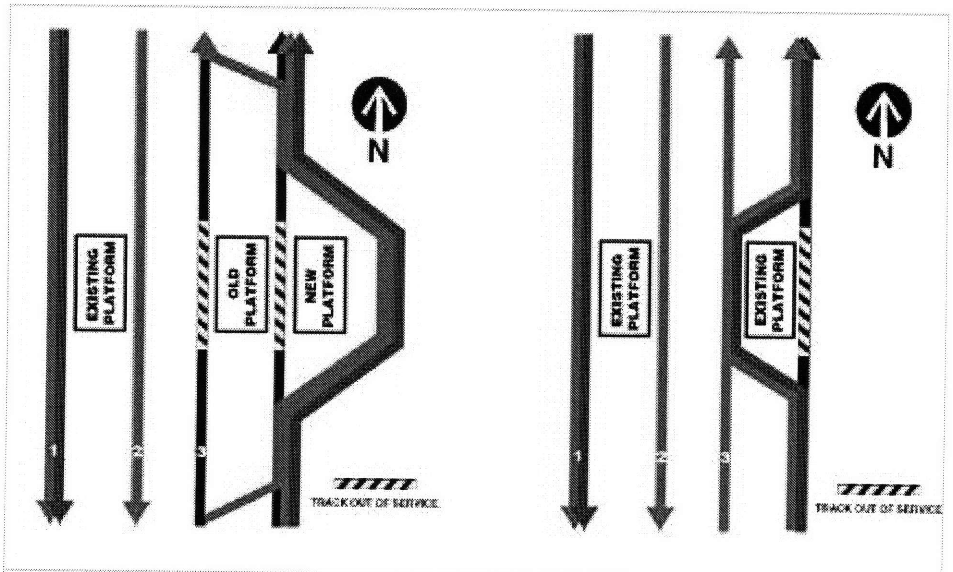


Figure 4: Configuration of Three track operation for Fullerton (left) and Belmont (right)
Source: CTA Three track operation brochure Spring 2007

An examination of the schedules for the Brown Line before the "Three track operation" shows that trains leaving the Kimball station at 7:00 A.M. (southbound) were expected to arrive to the Irving Park station at 8:03 A.M (northbound) after completing most of the "loop" cycle. On the other hand, train schedules for the same trip during Three track operations show that the same trip would arrive at 8:11 A.M at Irving Park. In summary, the project caused an 8 minute delay for every 63 minutes of total trip time during the AM peak hour.

The schedules also show a reduction in the number of scheduled trains during the AM peak. Between 7:00 AM and 8:02 AM, the Brown Line scheduled 20 trains under a 3 to 4 minute headway regime. During Three track operations, the CTA scheduled 17 trains serving a 4 minute headway. This means that, on average, customers waited 15% longer to board the train, and likely, under more crowded conditions.

As a response to the project-induced inconveniences, the CTA programmed additional bus service in the following routes:

- Morning rush period (6 a.m. to 9:30 a.m.): #11 Lincoln/Sedgwick, #22 Clark, #134 Stockton/LaSalle Express, #135 Clarendon /LaSalle Express, #151 Sheridan
- During the evening rush period (3 p.m. to 6:30 p.m.): #11 Lincoln/Sedgwick, #22 Clark, #147 Outer Drive Express, #148 Clarendon/Michigan Express28

As shown in Figure 3, these routes act like close substitutes for the Brown, Red and Purple lines by connecting the North of the city with the CBD

Apparently, the amount of additional service was enough to serve customers that decided to use the bus. Therefore, it was scaled back in half after May 17th,when president Ron Huberman announced a reduction in this supplementary service: *"The level of extra bus service being provided when three-track operation first began was costing nearly $150,000 a week. Although the CTA scaled back the extra service as people adjusted their commutes, adding back trains and reducing more of the supplemental bus service will allow us to bring that weekly cost down to just under $77,000 a week. When you add that up through 2009, a $73,000 per week savings is a major improvement for our bottom line."* [29]

3.4 Observed ridership impacts

Here we examine the ridership levels in the Brown Line during specific stages of the BLCEP. The measurements come directly from boarding counts at all the gates. The Automated Fare Collection (AFC) system and the CTA's intranet planning site provide a detailed database to account for observed impacts on ridership. The period between August of 2006 and July of 2007 is currently examined

3.4.1 Description of the analysis

[28] Three track operation brochure. Chicago Transit Authority, April 2007
[29] CTA press release 05/17/2007

During the period of study, some stations in the Brown Line were closed for reconstruction. This is the case of Kimball, in the north end of the line, and Francisco, the third station in the inbound direction. These stations were closed to the public on September 15th of 2006. Figure 3 presents the system map of the north side of Chicago.

Also, in December 2nd of 2007, the stations of Addison and Montrose were closed. In April 2nd the Southport station was closed as well. These events were expected to have immediate impact in ridership, since passengers that want to continue using the Brown Line will have to walk to their next closest station and might prefer to board a bus or simply stop using public transportation.

Another important event during this period of study is the operational restriction to trains in the North Side in April 2nd of 2007. In order to permit the construction of elevators in the Belmont and Fullerton stations, one of the two tracks that are usually shared by the Brown, Red and Purple Express lines in the outbound direction, was closed, so all three lines had to operate in only three tracks (two inbound, one outbound), causing slower trips and more crowded trains. This event was also expected to produce a ridership drop in the Brown Line.

The following analysis is divided in three parts. The first one will evaluate the overall ridership figures of the Brown Line by analyzing all the boardings during the period of study and comparing them to the expected number of boardings in that month, in order to understand the magnitude of the impact of the Brown Line's maintenance works in ridership. This analysis is done for weekday averages. In a similar way, the stations in the Loop are analyzed to see the extent of the ridership changes in the return trips.

The second part of the analysis consists of studying the boardings on each of the stations that were directly affected by the mentioned changes in service. This analysis also relies the comparison of boardings in each station against the expected number of boardings, in order to have a more detailed account of the ridership losses. It also evaluates the ramp-up period that follows the reopening of a rail station, in terms of how passengers return to use it after months of being inactive. This analysis is made on a day by day basis and it contributes to the understanding of the ramp-down period, as well that precedes the closing of a station.

A third analysis focuses on the changes in boarding times in the Brown Line and the Loop, in order to analyze if people decided to change their boarding times to avoid the resulting congestion. Changes in boarding times will also be analyzed by comparing them to last year's boarding time distributions

Table 8 presents a summary of the service disruptions during the period of study. The third column indicates the type of impact that each disruption had on the Brown Line customers should they had decided to continue using the Brown Line. The fourth column represents the changes in travel behavior that can be expected from customers. Although some of these disruptions are expected to have an impact on users of other rail lines, this first part of the analysis will be limited to examine boardings in the Brown Line and in the Loop as a natural trip generator for return trips.

Type of disruption	Affected Stations	Effects in the Brown Line travel experience	Expected effects in travel behavior
Station is closed	Kimball and Francisco (09/06), Montrose and Addison (12/06)	Longer walking times	Customers walk to next station, Customers use transit but switch to bus or Customers leave the system to use other modes
Station is reopened	Kimball (01/07), Francisco (03/07), Surrounding stations	Walking times return to status quo	Customers return to their home station, Customers continue using buses, Customers permanently abandon the system
Trains share tracks	All stations in the Brown Line and Red Line (04/07), Belmont and Fullerton (04/07)	Longer waiting times, Longer travel times, Crowded platforms	Customers use transit but switch to bus, Customers leave the system to walk or drive, Customers change their boarding times

Table 8 The Brown Line Capacity Expansion Project. September 2006-June 2007

3.4.2 Estimation of growth rates

In order to perform this analysis, a baseline was constructed to compare the actual recorded boardings vs. the expected boardings, which represent how much ridership would have grown –with respect to the previous month- if the project had not come into effect.

The calculation of the number of expected boardings required the estimation of the expected growth rate. This rate represents the cyclical monthly variations in ridership -caused by ongoing changes in conditions such as weather, economic activity and development- but exogenous to the maintenance-related impacts. Growth rates were calculated independently for both the Loop and the Brown Line. The estimation of these growth rates is essential because it allows us to separate the effects of season-induced changes in ridership from other effects, by comparing actual growth vs expected growth.

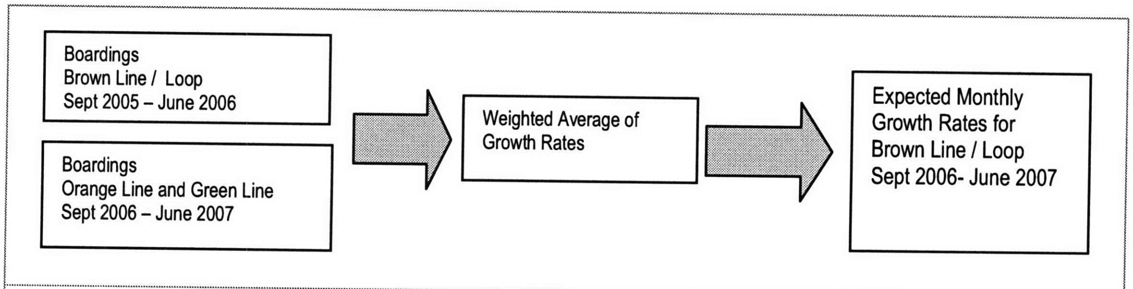


Figure 5: Analytical approach to estimating Monthly Growth Rates: Brown Line and the Loop

The estimated monthly growth rates are taken as the averages of the monthly Growth rates for the Green Line and the Orange Line (two areas of control) between 09-2006 and 07-2007, and the monthly Growth rate for the Brown Line (or the Loop) between 09-2005 and 07-2006. This average is weighted by the number of boardings on each line. Figure 5 illustrates the latter. This can be formulates as it follows.

The observed growth rate (G) for line *x*, in month *i* and year *y* is calculated with Equation 1, based on figures of weekday ridership (R).

$$G_{x,i,y} = \frac{R_{x,i+1,y} - R_{x,i,y}}{R_{x,i,y}}$$

Equation 1: Observed growth rate

And the expected monthly growth rate (EG) for line *a*, in month *i* and year *y*

$$EG_{a,i,y} = \frac{\sum_{x}^{control\_lines} G_{x,i,y} \times R_{x,i,y} + G_{a,i,y-1} \times R_{a,i,y-1}}{\sum_{x}^{control\_lines} R_{x,i,y} + R_{a,i,y-1}}$$

Equation 2: Expected monthly growth rate
Where
a is a line/section affected by the BLCEP [Brown Line, Loop]
x is a control line: [Orange Line, Green Line]

Table 9 shows the expected growth rates for the Brown Line and the Loop: Among the commonalities in the growth rates, it can be seen that September is a month when there is an important growth in ridership, mostly related to the start of activities that have been suspended during the summer such as schools, universities and other related services.

Also it is seen that ridership has an important drop in the month of December when assumingly some businesses cease activities and the winter weather forces some riders to use other modes.

	SEASONAL GROWTH RATES	
Line *i*	Brown Line Weekday	Loop Weekday
Sep '06	10.68%	8.16%
Oct '06	-2.58%	-1.43%
Nov '06	-4.26%	-5.13%
Dec '06	-7.76%	-8.24%
Jan '06	3.51%	3.41%
Feb '06	-0.16%	0.42%
Mar '06	3.59%	5.35%
Apr '06	0.31%	0.48%
May '06	4.83%	4.66%
Jun '06	-0.29%	1.62%
Jul '06	0.58%	5.22%

Table 9: Growth rates for Brown Line and Loop

Figure 6 shows the seasonal ridership cycle on an month-by-month basis. August has an indexed ridership of 1 –for the Brown Line and the Loop – and all other ridership figures depend on the estimated growth rates.

These growth rates represent the percentage of expected additional boardings on a particular month with respect to the boardings on the previous month. As mentioned before, these growth rates represent the seasonal ridership cycle. For instance, if the month of December recorded 1000 boardings in the Brown Line, then under normal conditions, 1035 boardings would be expected in the month of January.

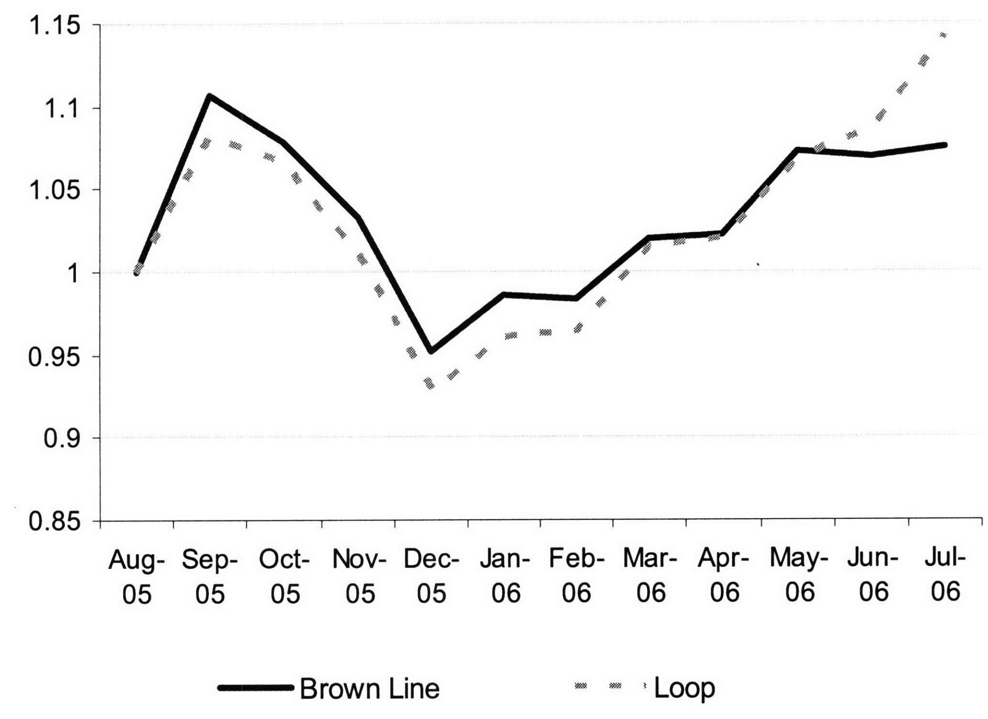


Figure 6: Estimated seasonal ridership cycle (indexed to August '05)

This approach has, however, at least two limitations: (1). The growth rates are calculated based on two other rail lines that serve as control zones. If the factors that induce transit ridership (density, land use, quality of service, etc) change at a very different rate between the control zone and the Brown Line, then the estimated growth rates will incur in a bias. (2). The growth rates are also calculated based on the ridership for the same rail line (Brown) in the same period for the year before. In the case of very different seasonal conditions between one year to the other, the estimation will also have a bias. One of the advantages of averaging the Lines of control in the current year with the Brown Line in the past year is that the potential effect of these biases can be partly ameliorated.

3.4.3 Changes in total boardings at the line level.

After estimating growth rates for the Brown Line and the Loop, it is possible now to calculate how many passengers would have boarded each line if there had been no disruptions in service.

Table 10 and Table 11 show how the average daily boardings changed in the months of interest for the Brown Line and the Loop. The comparison on a month by month basis is appropriate as a first approach to examine aggregate changes in travel behavior. The tables show that in the months where there were station closures, average weekday ridership in the Brown Line decreased between 3.4% and 5.1% with respect to the expectations; for the Loop, the area which should generate most of the return trips, there was a comparable decrease of 2.9% to 3.9% decline in ridership with respect to the expectations.

Brown Line - weekday averages	Actual	Expected	Loss/gain	% difference
August '06	46736	46736	-	-
September '06 *	50002	51726	-1724	**-3.45%**
October '06	49233	48714	519	1.05%
November '06	47056	47137	-81	-0.17%
December '06 *	41283	43407	-2124	**-5.14%**
January '06	43891	42733	1158	2.64%
February '06	44228	43822	406	0.92%
March '06	46199	45816	383	0.83%
April '06 *	40801	46344	-5543	**-13.59%**
May '06	43092	42773	319	0.74%
June '06	43188	42965	223	0.52%

Table 10: **Actual vs. Expected ridership – Brown Line**
*** Months where there was a disruption in service**

These tables also provide an idea of the forecast's accuracy. In those months when there was no disruption, there was an average difference of only 418 more passengers with respect to the forecasted ridership. This is equivalent to an average 0.9% difference from the expected forecast in the Brown Line, with a standard deviation of 0.8%.

Loop - weekday averages	Actual	Expected	Loss/gain	% difference
August '06	67491	67491	-	-
September '06 *	70201	72996	-2795	**-3.98%**
October '06	67778	69195	-1417	-2.09%
November '06	67075	64303	2772	4.13%
December '06 *	59833	61548	-1715	**-2.87%**
January '06	62649	61874	775	1.24%
February '06	62463	62915	-452	-0.72%
March '06	65751	65807	-56	-0.09%
April '06 *	64216	66068	-1852	**-2.88%**
May '06	67760	67211	549	0.81%
June '06	69681	68858	823	1.18%

Table 11: Actual vs. Expected ridership – Loop
*** Months where there was a disruption in service**

However, the picture is mixed when examining the Loop figures: There is an average difference of 427 daily passengers with respect to forecasts. This is equivalent to an average 0.64 % difference from the expected forecast in the Brown Line with a standard deviation of 1.94% (more than three times the size of the difference). This larger inaccuracy can be explained because the Loop is connected to all the other rail lines of the system. Most of these other lines are not directly affected by the events of the BLCEP. Also, the two control zones that were used are lines that cross areas that are natural trip generators, as opposed to the Loop's nature as Chicago's trip attractor

The most significant effect on ridership was on April, when more than 13% of the total Brown Line ridership declined, with respect to expected figures. On this month, operations in the Brown Line changed because one of the outbound tracks was closed due to construction work (three track operations). Surprisingly, the Loop figures do not show such a steep decline. Although there is a decline of almost 3% with respect to expectations, the Loop is the destination of most of the inbound trips and thus, the return trips should have presumably been affected in a similar way. Even in absolute numbers, there is no correspondence. The Brown Line had a reduction of more than 5,500 boardings while the Loop only had a reduction of 1,800 boardings. There are many possible explanations to this question, but ultimately, only an individual passenger-based analysis may provide the answer. Nevertheless, at least two hypotheses can explain this: (1) There could be have been a steep increase in rail users that board other lines on the Loop and thus, offset the three track operation related losses. (2) Passengers could have been less inclined to take a bus in the return trip; instead, the majority of the customers changed their boarding time in order to avoid congestion, rather than of changing their mode.

3.4.4 Changes in total boardings at the station level

This section analyzes the impact that station closures and station re-openings had on boardings. Average weekday figures from each station and their surrounding counterparts are studied.. Also, daily boardings are studied on the days before each of the station closes and after it reopens.

3.4.4.1 Station closures

(A) September 2006

On September 15th of 2006, both the Kimball and Francisco stations were closed. This event led a number of passengers to abandon the Brown Line or to use the nearest open station. In the case of the Kimball station, passengers could presumably walk to Kedzie, while in the case of Francisco, they could walk to Kedzie or Rockwell. As can be seen in Table 12, both Kedzie and Rockwell had a significant increase in their average daily ridership. Most likely the cause of this increase in ridership was the closure of Kimball and Francisco stations.

Station**		Sep 15-Oct 15 / 2006	Increase in rides
1. Kimball	Actual	0	Closed
	Expected *	3837	
2. Kedzie	Actual	4794	~ 300%
	Expected *	1606	
3. Francisco	Actual	0	Closed
	Expected *	1029	
4. Rockwell	Actual	1708	~ 17%
	Expected *	1457	
Total 4 stations	**Actual**	**6502**	**N/A**
	Expected	**7929**	

Table 12: Actual vs. Expected weekday ridership - Station Closures Kimball and Francisco
* Expected ridership in month (i) = ridership in month (i-1) x brown line growth rate month (i)
** Distances between stations: 1-2 = 411 m ; 2-3 = 601 m ; 3-4 = 630 m

Kimball and Francisco together were expected to record around 4,900 average daily boardings between September 15th and October 15th. However, some of these passengers left the Brown line in favor of other ways to commute. Other passengers went to the nearby Kedzie and Rockwell stations, which recorded an additional 3,450 average daily boardings. Out of these additional boardings, less than 300 were recorded in Rockwell, leaving the majority of these diverted passengers to Kedzie (an increase of

300%). However, at this point of the analysis, it is impossible to know how many of these new Kedzie boardings belong to former Kimball or Francisco customers. Finally, 1,450 weekday boardings did not continue using the Brown Line, representing a 31% of the expected boardings in Kimball and Francisco together.

Figure 7 shows the boardings in the immediate weekdays before the station closure. On one hand, it can be seen that both in Kimball and in Francisco, the ridership numbers are low with respect to expectations. The week before the closures, Kimball and Francisco recorded an average of 3,938 and 1,058 weekday riders respectively: 8.3% and 24% below the expectation. Similarly, two weeks before the closures, these two stations recorded an average of 4,058 and 1,085 weekday riders respectively: 5.5% and 22% below expectations.

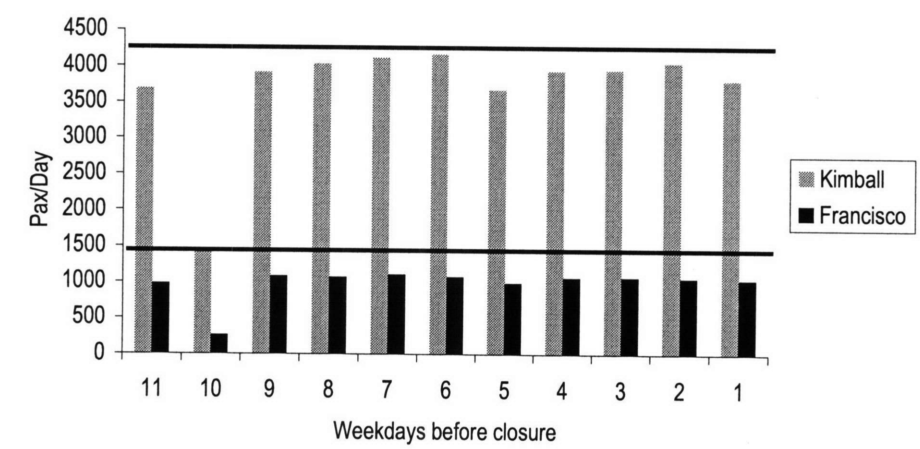


Figure 7: Weekday ridership before station closures – Kimball and Francisco
* Lines represent the respective expected ridership figure for each station in the month of September.

These figures suggest the following: (A)- Losses in expected ridership appear with anticipation. This is explainable because the project was announced with considerable anticipation and there was a direct effort to communicate the particular dates when a station would be closed. (B)- The process of leaving the system is gradual. The ridership figures are slightly lower on the immediate week before the closures than on the previous week. Although the ridership differences from one week to the next are very low, it could be argued that there is a slight ramp-down effect which can explain how passengers systematically leave the system before an announced disruption is about to happen. However, the evidence is disputable as the observed difference in weekly ridership could be explained by a heterogeneous expected level of ridership across weeks of the same month, which is a factor unrelated to the station closure.

(B) December 2006

On December 2nd of 2006, the Montrose and Addison stations were closed to the public. Just like in the case of Kimball and Francisco, the nearby stations also registered an increase in boardings, which can presumably be attributed to diverted passengers from the stations that closed. Montrose and Addison were expected to record 3,700 average daily boardings in the month of December. Some of those riders were diverted to the nearby stations of Damen, Irving Park and Paulina. These stations, altogether, recorded 2,000 new passengers. This increase in customers can be mostly attributed to former Montrose or Addison passengers. Finally, there were 1,700 weekday boardings that did not continue using the Brown Line. This represents 46% of the total expected boardings on Montrose and Addison.

Station**		Dec / 2006	Increase in rides
1. Damen	Actual	2431	~ 35%
	Expected*	1802	
2. Montrose	Actual	0	Closed
	Expected*	1967	
3. Irving Park	Actual	2930	~28 %
	Expected*	2296	
4. Addison	Actual	0	Closed
	Expected*	1724	
5. Paulina	Actual	2765	~ 38%
	Expected*	2008	
Total 5 stations	**Actual**	**8126**	**N/A**
	Expected	**9797**	

Table 13 : Actual vs. Expected - Station Closures Montrose and Addison
* Expected ridership in month (i) = ridership in month (i-1) x brown line growth rate month (i)
** Distances between stations: 1-2 =570 m ; 2-3 = 795 m ; 3-4 = 840 m ; 4-5 = 491 m

It is also worth noting how passengers apparently decided what station to board once their home station was closed. Damen received 630 new boardings, Irving Park received 630 and Paulina 760. Although geographically, Irving Park was the station in the middle –and thus- should have received the most boardings in total, surprisingly this was not the case. Factors inherent to the new travel experience, such as additional walking time, adequate walking conditions and station design may have influenced passengers' choice. For instance, the distance between Irving Park and its closed neighboring stations is 795m and 840m, while the distances from these closed stations to Damen and Paulina is 570m and 491m respectively.

3.4.4.2 Station reopening

Two station reopened in the period of study, Kimball and Francisco. The reopening dates were January 12th and March 9th respectively. The reopenings were expected to return most passengers to their original home station, but in a gradual way. In a similar fashion, some passengers might not return to the Brown Line because they have used another mode or because they have changed their destination for reasons unrelated to the station closures.

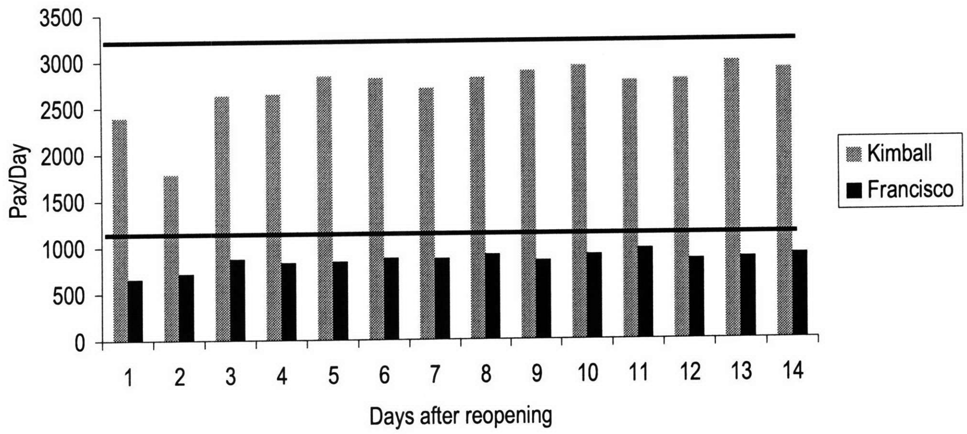


Figure 8: Ridership after station re-opening – Kimball and Francisco

Figure 8 shows the boardings in the immediate weekdays after the stations were reopened. As it can be seen, there is a significant gap between post-reopening daily ridership and expectations. Kimball, the first week after reopening, recorded an average of 2,460 daily boardings, while the second week recorded average, 2,830 daily boardings, while the expected weekday ridership for January 2007 was 3,363 daily boardings. Similarly, Francisco recorded 790 and 880 average weekday riders on the first two weeks after reopening respectively. However, the expected number of average rides for Francisco, in March of 2007 is 1,043 daily rides.

These observations show that even two weeks after reopening the stations, ridership levels did not return to their expected levels. Francisco experienced only 85% of the expected ridership after the second week, and in Kimball, this figure is 84%. Presumably, the missing passengers changed their travel patterns after having their station closed by a moderate period of time. An individual analysis of the passenger's behavior can confirm this hypothesis.

However, there is no sufficient evidence to affirm that these passenger losses are permanent. First, the window of observation is only two weeks after the station reopening, which means that in further weeks there might be a return of passengers. Second, BLCEP is a broader project which has more instances than the mere station closures. Only when the project is completed will it be a case for a compelling analysis of long term losses.

3.4.4.3 Three track operation – April 2007

On April 2nd of 2007, one of the four tracks that serve the Brown, Red and Purple Lines was closed due to works in the Belmont and Fullerton stations. This track closure had effects on train operations, as fewer trains were scheduled (with longer headways) and customers faced frequent delays on their trips. As expected, ridership in the Brown Line decreased significantly. On average, there was an 11% passenger loss per station, ranging from 3% to 20%. The only station where ridership grew in this month was Paulina, likely caused by the simultaneous closure of the Southport station on April 2nd.

Station	Apr-2006		Effect of 3 Track	
	actual	expected	Pax. loss	% dif
Kimball	3189	3292	-103	-3.2%
Kedzie	1738	2080	-342	-19.7%
Francisco	916	1053	-137	-13.0%
Rockwell	1376	1567	-191	-13.9%
Western	3112	3342	-230	-7.4%
Damen	2369	2665	-296	-12.5%
Montrose	closed		-	-
Irving Park	2881	3182	-301	-10.4%
Addison	closed		-	-
Paulina	3425	3015	+410	+12.0%
Southport	closed		-	-
Wellington	2337	2547	-210	-9.0%
Diversey	3893	4426	-533	-13.7%
Armitage	3301	3764	-463	-14.0%
Sedgwick	2458	2806	-348	-14.2%
Chicago	4303	4642	-339	-7.9%
Merchan Mart	5501	5678	-177	-3.2%

Table 14: Three track operations: Impact on Brown Line ridership

Table 14 presents the Brown Line stations organized according to their location on the Line, where Kimball is the farthest stop from the Loop. Indeed, distance to the loop does not seem to be a factor directly related to boarding losses. On one hand, walking can be a good substitute to short trips. On the other hand, long trips will be more affected by delays and slower trains, and thus, these passengers may be more inclined to leave the Brown Line. Also, there are many other factors that will influence a passenger's choice such as the availability of other modes and his/her personal social and economic background.

3.4.5 Changes in boarding times

Another potential change in travel behavior is related to the changes in boarding times, indicating efforts to avoid induced system crowding in the peak hour, or simply because their trip took longer and there is a natural need to board earlier. This change can be detected between March and April, when the three track operation started. Since peakspreading is one way to ameliorate congestion, it is important to study how passengers reacted in the particular case of the Brown Line.

March'07	Kimball	Kedzie	Francisco	Rockwell	Western	IrvPark	Paulina	AVERAGE
6:30-7:00	14.81%	11.94%	7.77%	8.42%	8.65%	8.17%	7.74%	9.64%
7:00-730	22.47%	22.04%	18.45%	18.81%	20.15%	18.32%	16.36%	19.51%
7:30-8:00	22.06%	22.48%	27.18%	24.65%	25.15%	25.72%	27.10%	24.91%
8:00-8:30	18.52%	21.28%	30.10%	28.32%	25.34%	27.21%	28.04%	25.54%
8:30-9:00	12.02%	12.49%	11.65%	12.57%	12.67%	13.60%	14.60%	12.80%
9:00-9:30	10.12%	9.77%	4.85%	7.23%	8.03%	6.98%	6.16%	7.59%

Table 15: Distribution of boardings in the peak hour per station –March, Brown Line

April '07	Kimball	Kedzie	Francisco	Rockwell	Western	IrvPark	Paulina	AVERAGE
6:30-7:00	16.81%	14.03%	8.91%	10.67%	10.11%	9.25%	9.19%	11.28%
7:00-730	21.68%	21.80%	21.29%	19.81%	21.04%	19.83%	18.59%	20.58%
7:30-8:00	22.12%	21.80%	25.25%	25.44%	25.89%	26.08%	27.52%	24.87%
8:00-8:30	17.17%	19.89%	27.72%	24.50%	21.79%	24.95%	25.35%	23.05%
8:30-9:00	11.42%	12.26%	11.39%	12.54%	12.77%	12.57%	13.01%	12.28%
9:00-9:30	10.80%	10.22%	5.45%	7.03%	8.40%	7.32%	6.35%	7.94%

Table 16: Distribution of boardings in the peak hour per station –April, Brown Line

Table 15 and Table 16 show the percentage of trips that boarded on each half-hour time segment within the peak hour, for the months of March and April. For instance, from all the trips that boarded at Rockwell station in the month of April between 6:30 and 9:30 a.m., only 7% of them boarded in the 9:00-9:30 a.m. time segment. Those stations that are in the vicinity of the Southport station were not included because this station was closed at the same time and including them could distort the analysis The selected stations were sorted according to their location on the Line, where Kimball is the farthest stop from the Loop. It can be seen that stations that are located farther from downtown have, on average, earlier boarding times, as expected. For instance, Kedzie recorded in April 14% of its AM peak boardings between 6:30 and 7:00, while Paulina only recorded 9%. This is expected because the Loop attracts most of the jobs and those who live farther from downtown will have to start their work trip earlier.

Mar to Apr	Kimball	Kedzie	Francisco	Rockwell	Western	IrvPark	Paulina	Average
6:30-7:00	2.00%	2.09%	1.14%	2.25%	1.46%	1.07%	1.45%	1.64%
7:00-730	-0.79%	-0.24%	2.84%	1.00%	0.89%	1.51%	2.22%	1.06%
7:30-8:00	0.07%	-0.68%	-1.94%	0.79%	0.73%	0.37%	0.42%	-0.03%
8:00-8:30	-1.35%	-1.39%	-2.37%	-3.82%	-3.55%	-2.26%	-2.69%	-2.49%
8:30-9:00	-0.60%	-0.22%	-0.26%	-0.03%	0.10%	-1.03%	-1.59%	-0.52%
9:00-9:30	0.67%	0.45%	0.59%	-0.19%	0.37%	0.34%	0.19%	0.34%

Table 17: Change in percentage of peak hour boardings per ½ hour segment between March and April. Brown Line stations

Table 17 shows how the boarding times shifted between March and April. Although small in percentages, passengers tended to start boarding earlier. On one hand, the first two segments in the morning increased their share while the segment between 8 and 9 a.m. was the one that had the larger loss. On the other hand, there was a small percent increase of passengers boarding later, perhaps representing those commuters that have a more flexible schedule and take advantage of it.

3.5 Conclusions

The analyses of the total boardings at the line level and at the station level show a change in travel patterns during the different stages of the Brown line capacity expansion project. As stations were closed to the public, the nearby stations on the same rail line recorded increases in boardings, suggesting that some passengers were willing to walk longer distances to board those trains rather than switching to the bus. The available evidence shows a net loss, however, in total station boardings, suggesting some passengers may have switched modes or changed their travel activities.

On the other hand, it is very revealing to notice that once the examined stations were reopened, fewer passengers than expected returned to the station. This means that some passengers did not retake their same travel habits inmediately after two weeks of examination in two Brown Line stations. Changes, however, could have also been triggered by reasons not related to the BLCEP. Finally, evidence shows that the distribution in boarding times changed as soon as the three track operation started. A higher proportion of the passengers decided to board at earlier times, suggesting that passengers were forced to start their trips earlier due to the longer trip times or to avoid the resulting crowding.

The next chapters of this thesis focus on the analysis of individual passengers' behavior instead of changes in recorded boardings. The records of the Smart Cards allow us to infer the particular choices made by each passenger and understand better the motivating factors that can influence an individual's decision to select a commuting mode. This type of work entails other methodological challenges as described in the next chapter.

4 Methodology to extract travel information from Smart Cards

This chapter shows how behavior of transit users can be examined through the available data sources at the CTA. This research draws data from the AFC and AVL systems, the active Smart Cards list and GIS maps from Chicago. This chapter describes the methodology in order to point its limitations and improve the likelihood of being replicated for future research.

While this thesis analyzes how individual passengers modified their travel behavior once the maintenance project started, this chapter is limited to explaining how they behaved *before* the project started. Once a behavioral baseline is set for each passenger, then in Chapter 5 we examine the changes he/she made *during* some time after project (station closure, three track operation) or *after* the project ended (station is reopened).

An important caveat is that the focus of this research is restricted to commuters. The reason for this limitation is that this is a group of passengers for whom it is possible to infer the destination of their trips with a reasonable level of confidence. Since the CTA does not record when a passenger leaves the system, this research infers the destination of a commute trip based on the examination of the location of the afternoon return boardings. Section 4.3.1 explains in detail the methodology used. In the case of non work trips or non-home based work trips, this becomes a more challenging task. Furthermore, examining commuters is important because they represent a large share of the peak hour demand and understanding their reactions to a maintenance project will offer valuable insights for developing critical tasks such as planning an adequate level of supplementary bus service and estimating the project-related costs and revenue impacts.

4.1 Data sources

This study draws data from various sources and links them together to provide a good representation of a passenger's behavior. The following sections explain the characteristics of these main sources and the drawbacks of this methodology.

4.1.1 List of Smart cards

The Chicago Card and Chicago Card Plus programs require the user to submit an application form in order to receive the physical card. This application contains basic information about the client such as his name, billing address (optional) and phone number (See Table 18). These data are kept by the contractor who manages the Chicago Card and Chicago Card Plus programs, but can be shared with the CTA by request for planning purposes. The most important item from this list is the unique ID that identifies each card. This is specified by the card's manufacturer and is used to identify the transactions of each card owner. The complete list of active Smart Cards with a reported address included 196,615 records as of December of 2006.

DATA FIELD	MEANING
AFC_NO	Serial Number ID of the Smart Card
ADDRESS	From reported billing address
CITY	From reported billing address
STATE	From reported billing address
ZIP	From reported billing address

Table 18: Smart Cards data table

The reported addresses were geo-coded using the software TransCad[30]. This tool is able to assign longitude and latitude coordinates to each to the cards based on specially coded street maps with street names and numbers, as well as zip codes and municipality names. Unfortunately, some of these addresses were incorrectly reported or reported outside the Chicago Metropolitan Area. This caused a reduction in the number of available cards for this study to 158,708, a loss rate of 19.2% due to these inaccuracies.

[30] Transcad manual, Caliper corporation

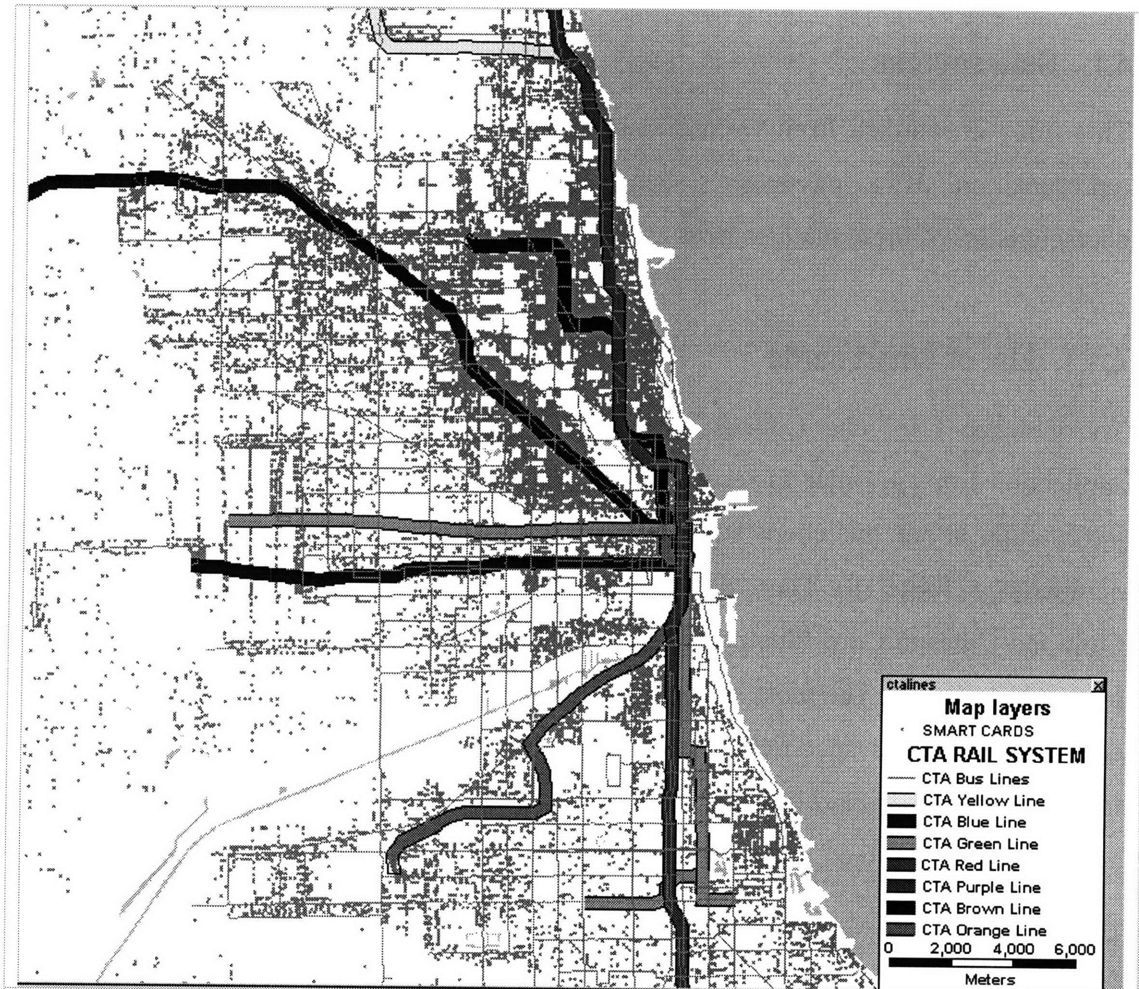


Figure 9: Map with geo-coded Chicago Cards

Figure 9 shows a density dot map with the spatial distribution of the cards in the City of Chicago. As can be seen, there is a higher density of cards in the north of the city, consistent with recent research made by the CTA (2007)[31]. Most of the cards are clustered around the rail lines, while some areas only served by bus are practically empty.

4.1.2 Automatic Fare Collection system:

The AFC system which was inaugurated in the CTA in the year 1999, records all the transactions in rail stations and buses and produces a detailed list of all the daily

[31] CTA, Market research.(2007) CTA Smart Card market: What do we know?.

boardings. In the case of the rail system, the station turnstiles are continuously sending data to a central server. In the case of bus fare boxes, the recorded data are downloaded on a daily basis by the driver when the bus enters the depot.

Table 19 shows the type of data that the AFC system keeps. As can be seen, when a transaction is made with a Smart Card, the system keeps track of the individual ID of each card.

DATA FIELD	MEANING
SEQNUM	Consecutive number of each transaction
SER_NBR	Serial Number ID of the Smart Card
USE_LLT	Date and time of transaction HH/MM/DD _ hh_mm_ss
EQUIP_ID	ID of the entrance/farebox where the transaction ocurred
TRANS_EVENT	Type of transaction
CURRENT_ROUTE	ID of the bus route
LAST_ROUTE	ID of the last bus route
FARE_MEDIA_TYPE	Type of Fare Media
ENTRANCEID	ID of the entrance
STA	ID of the rail station
BUS_FLAG	1 of this is a bus trip
TRIP_FLAG	1 if this is a trip
BUS_NBR	ID of the Bus

Table 19: AFC data table

4.1.3 Automatic Vehicle Location (AVL) system:

The AVL system tracks all the buses that operate in the CTA network while they are in operation. Buses are equipped with a dead-reckoning enhanced GPS system which identifies the location of the vehicle as it makes progress in the route. Similarly to the AFC system, all the AVL system's records are downloaded and kept in a single database. The system records a time-stamp every time it approaches or leaves a bus stop and associates it with a X-Y coordinates. Table 20 shows some of the fields that comprise the resulting AVL records.

DATA FIELD	MEANING
EVENT_TIME	Time of the signal
BUS_ID	ID of the bus
ROUTE_ID	ID of the bus route
DIRECTION	Direction of the bus on the event time
GEOID	ID of the bus stop type I
STOPID	ID of the bus stop type II
TAGEOID	ID of the bus stop type III
GEODESCRIPTION	Address of the nearest bus stop

Table 20: AVL data table

4.1.4 Linking AFC with AVL data

For the purposes of this study, it is important to determine the boarding locations of each passenger. In the case of rail trips, the AFC system records the train station where a passenger enters the system. The level of the data is detailed enough so that it is possible to know if the transaction was made in a particular entrance or even on a particular turnstile. However, the AFC system does not provide this level of detail for bus trips, because the bus stop location is not automatically recorded. Nevertheless, this information can be found on the AVL dataset.

In order to link both databases, a procedure, initially described by Zhao (2004)[32] and later documented by Gupta (2006)[33] was used by CTA staff to provide the source data for this research. This procedure has the following logic: a-) Match the BUS_NBR (AFC) and BUS_ID (AVL) fields for a particular customer to link the identification that each bus has on both systems b-) Match the USE_LLT (AFC) and EVENT_TIME (AVL) fields for a window of five minutes, to link a passenger's boarding time with the arrival of a bus to a stop. c-) Match the geographical location of a bus at the transaction time with the nearest bus stop's coordinates. The critical step of this process is to match the times of the AFC transactions with the time-stamps of the AVL when it arrives-departs a bus stop. This process, however, has a high rate of matching, showing that up to 99.1% of the AFC records are matched within 5 minutes of the AVL time.

[32] Zhao, Jinhua (2004) The Planning and Analysis Implications of Automated Data Collection Systems: Rail Transit OD Matrix Inference and Path Choice Modeling Example. MIT
[33] Gupta, Saumya (2006) Understanding Transit Travel Behavior: Value added by Smart Cards. MIT thesis

This procedure permits the study of all the transactions that were made on a particular day by the geographical location of the boarding.

4.1.5 Select monthly datasets

Monthly datasets were extracted for further study. These datasets had the following characteristics:

- There is one dataset for each month between (and including) September of 2006 and May 2007, hence, a total of nine datasets were downloaded
- Each dataset comprises a seven day consecutive period without extraordinary weather events
- Each dataset is separated into at least 3 weeks from the next one and 5 weeks at most.
- Each dataset includes all the transactions made with Chicago Cards and Chicago Card Plus for the selected weeks.

Table 21 shows the total amount of records extracted for each month. As mentioned, this total represents a period of 7 straight days within each month. It can be seen that around 72% of all transactions that happen each day are boardings. This number stays fairly constant across months, just like the percentage of transfers which is approximately 19% in all the system. A remaining percentage of transactions (between 9 and 10%) corresponds to those transactions that did not result in a boarding, like adding value to a card, requesting information for the stored value, or aborting a transaction.

Month	Days recorded	Total transactions	% of which are boardings	% of which are transfers	Average smart card weekday boardings
September 2006	01-07	1,382,864	71.4 %	19.1%	286,554
October 2006	04-10	1,539,624	72.4 %	18.9%	278,261
November 2006	01-07	1,555,602	72.7%	18.7%	275,505
December 2006	03-09	1,517,962	72.8%	18.9%	268,452
January 2007	06-12	1,543,619	72.8%	18.9%	273,952
February 2007	08-14	1,432,346	73.3%	19.0%	259,340
March 2007	03-09	1,471,499	73.1%	19.0%	265,490
April 2007	04-10	1,394,992	72.7%	19.1%	260,658
May 2007	01-07	1,494,013	73.1%	18.8%	264,552

Table 21: Extracted records per month

The last column of the table represents the average weekday boardings made with smart cards. These figures are extracted directly from each month's database, but they are representative only for the week that was selected. These monthly values resemble the seasonal effects on total system boardings, shown in Figure 10. As seen, the total monthly system ridership rises and drops in the system more pronouncedly than Smart Card-only rides. This could be explained under the hypothesis that frequent users are more likely to get Smart Cards because of the initial cost incurred in acquiring one the card. Therefore, tourists and occasional CTA users, which represent an important part of the seasonal changes, would not be as highly represented in Smart Card samples than commuters and transit dependent residents.

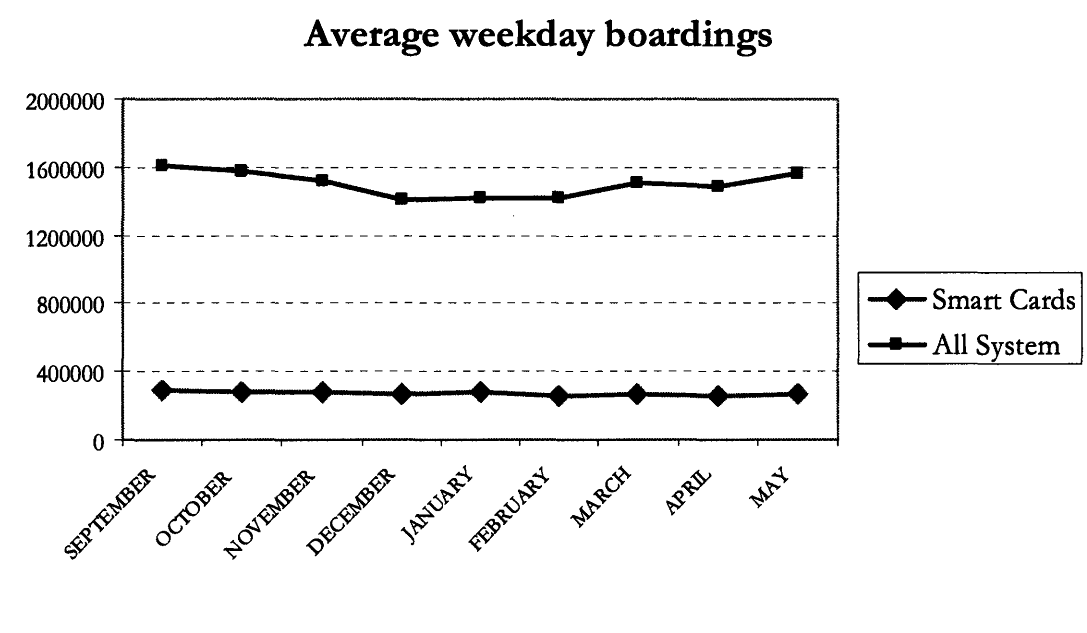


Figure 10: Weekday boardings: Smart Cards vs. all fare media

The next stages of the data selection process are station specific, meaning that the total sample will be narrowed to customers affected by the maintenance project. Figure 11 shows the logic of the filtering process. The final goal of this process is to identify the boarding patterns of a sample of cards which represent passengers that live near to a station and had a reduction in its level of service. The boarding patterns will ultimately be described by the following variables:

- Consistency of use (trips per week)
- Mode (% of bus trips and % of rail trips in the week)
- Boarding time (within the A.M. peak)
- Boarding location

Each step of the filtering process reduces the size of the sample, thus it is important to realize the type of information which is being removed. Just like the comparison of total boardings vs. only smart card boardings, a brief analysis will be presented on the type of results of each filtering step.

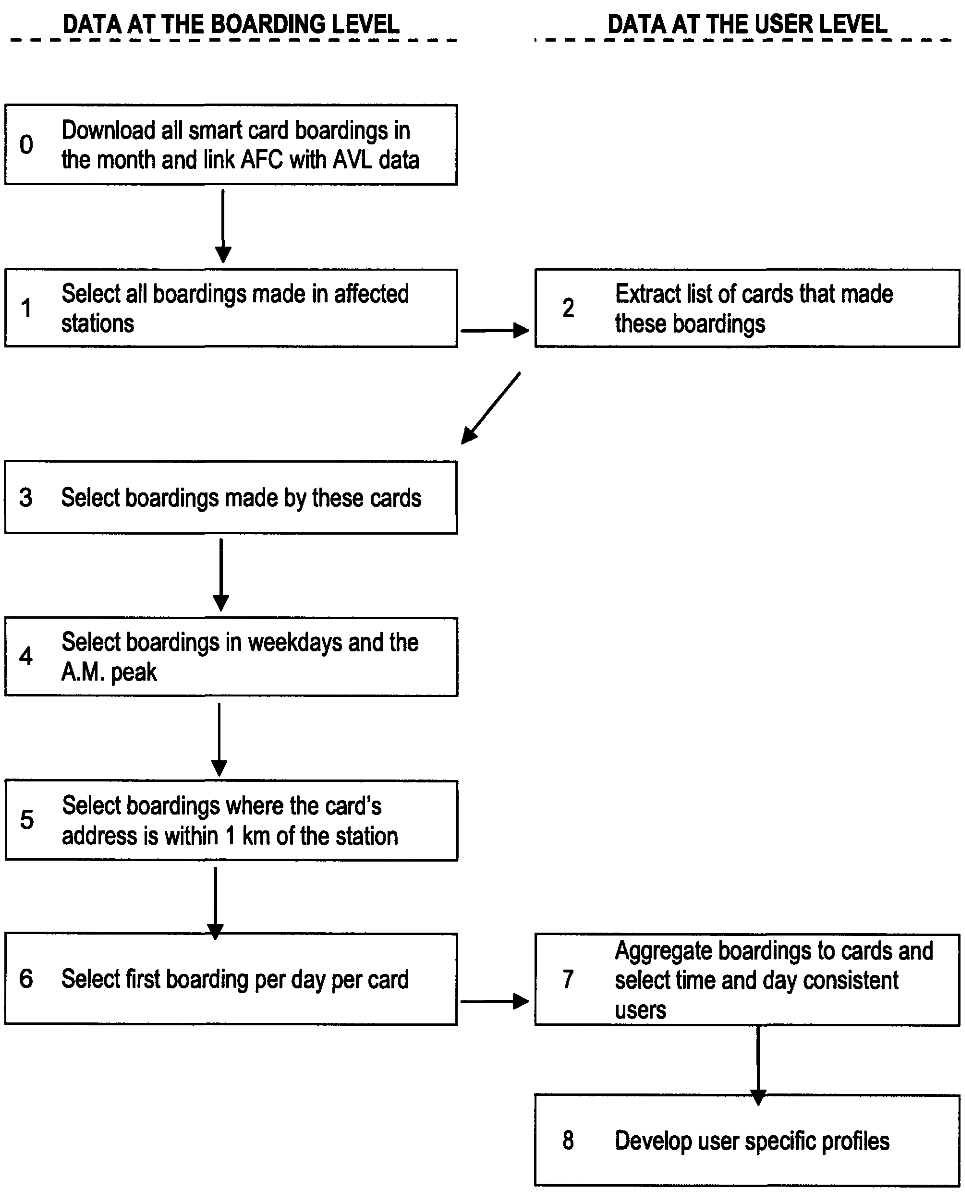


Figure 11: Data manipulation process for AFC-AVL linked records in the AM peak before the project

4.2 Morning peak patterns

In order to describe commute patterns, the first set of boardings to be examined is the morning boardings *before* the project took place. The idea is to set a baseline of travel behavior to be compared against morning boardings *during* the project.

4.2.1 Step 1 and 2: Select cards that use affected stations *before* the project

Stations are selected based on the timing of a project. Four stations were closed between September and December of 2006 (Francisco, Kimball, Montrose, Addison), as shown in Table 22, hence, these 4 stations were selected. Also, starting in April 4th, Three track operation affected all the commuters using the Brown, Red or Purple Express lines. In this case, only 16 stations were selected for analysis in order to reduce the burden in data processing

Step 1 consists in selecting all the *boardings* made in these stations from the respective datasets before the service disruption. In this case, we used the datasets of September-06 and November-06 because these correspond to all Smart Card activity 2 weeks before the closures. Also, we used the dataset of March-07 as it corresponds to activity 3 weeks before Three track operation. Step 2 consists in creating a list of all the *cards* that did these boardings. Table 22 presents a summary of the locations and the resulting number of cards that were selected before each service type of disruption.

Dataset	Type of disruption	Number of Cards	Selected Stations	
September-06	Station Closure	2,396	Francisco Kimball	
November-06	Station Closure	3,531	Montrose Addison	
March-07	Three track operations	39,749	Loyola Bryn Mawr Wilson Sheridan Argyle Davis Main Linden	Western Irving Park Paulina Belmont Wellington Diversey Armitage Addison

Table 22: Number of cards selected in the months *before* a service disruption

4.2.2 Step 3: Select all the boardings made by cards

With the list of cards that results from Step 2, a selection is made from each of the datasets to extract all the boardings made by the previously listed cards, regardless of where they were made. In this way, it is possible to monitor all the travel activity of these users before the maintenance project came into play. Table 23 shows the number of boardings recorded by the selected cards.

Dataset	Type of disruption	Number of Cards (N = 45,676)	Number of Boardings (N=432,136)
September	Station Closure	2,396	24,150
November	Station Closure	3,531	27,316
March	Three track operations	39,749	380,670

Table 23: Number of cards and boardings selected in the months before service disruptions

As can be seen, the trip generation rate varies for each month. In September, the records show 10.1 trips per week as opposed to only 7.7 trips per week in November. In March, this number increases again to 9.6. It is not the purpose of this study to explain the different trip rates, however this differential is consistent with the distribution of seasonal effects found in section 3.4.2, suggesting that higher monthly ridership can be related to different trip generation rates.

4.2.3 Step 4: Select by day and time

Step 4 consists in filtering the boardings by day of the week *and* time of day. Only weekdays were selected; the A.M. peak was selected to track the first trip of all users; the A.M. peak was broadly defined between 6 AM and 10 AM in order to capture as many users as possible.

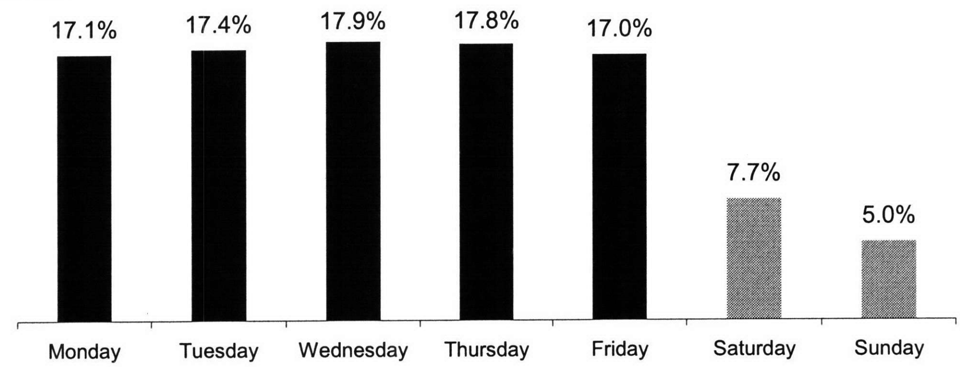


Figure 12: Boardings by day of the week (N=432,136)

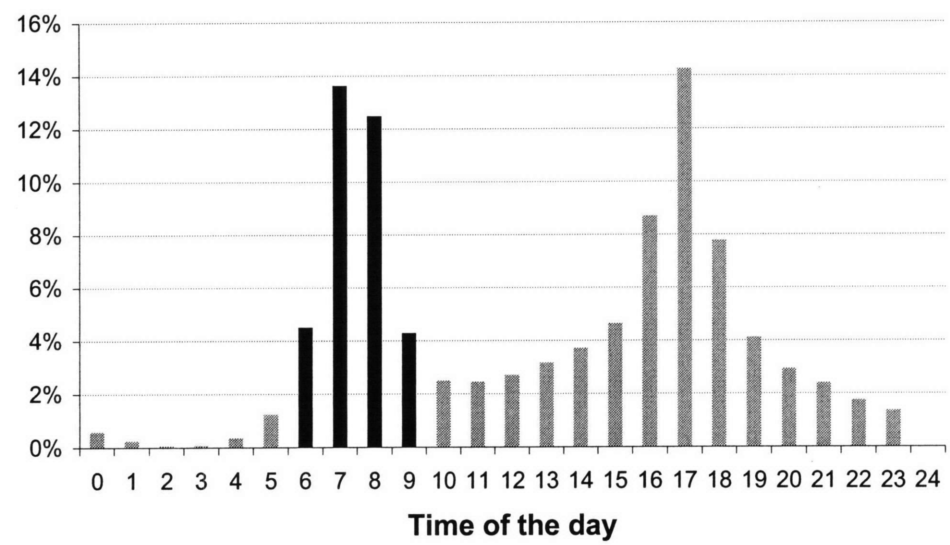


Figure 13: Boardings by time of the day (N=432,136)

Figure 12 and Figure 13 show the time and day distribution for the boardings. This research is interested those boardings in the black bars. After selecting these specific boardings, the sample was trimmed to 129,606 boardings distributed as shown below:

Dataset	Number of Boardings (N=129,606)
September	7,381
November	8,245
March	113,980

4.2.4 Step 5: Select cards with a reasonable address

A next step in the filtering process requires the identification of those cards which reported the address in the registration form and which complied with the described boarding patterns. Because the customer can write down any mailing address, or choose not write any, it is important to select only those boardings made where the registered address is near the boarding rail station or bus stop.

First, a filter eliminated all boardings from cards that do not have any registered address or that are not in the initial list of cards described in section 4.1.1. From the 129,606 boardings that resulted from the time of day and day of week filters, only 68,938 met the above mentioned criteria. This means that up to 47% of the boardings were eliminated from this analysis because of a low rate of address registration.

Second, a new selection was made from the 68,938 boardings that had a card with a registered address. This new criteria involved the aerial distance between the boarding station/stop and the address which was calculated based on the reported coordinates using Equation 3:

$$D(km) = k \times \sqrt{(([CC\_X] - [BRD\_X])^2 + ([CC\_Y] - [BRD\_Y])^2)}$$

Where:

K = 113.325 is the constant that converts distances from XY coordinates to kilometers

CC_X is the X coordinates of the reported address

CC_Y is the Y coordinates of the reported address

BRD_X is the X coordinates of the boarding station/stop

BRD_Y is the Y coordinates of the boarding station/stop

D(km) is the distance between two points in kilometers

Equation 3: Aerial distance between reported address and boarding location

After calculating all the respective distances to the boarding locations, only those boardings with a distance smaller than 1 km were selected. This is equivalent to creating a 1 km buffer around each boarding location in order to maximize the chances that the reported address is, indeed, the current household. Figure 14 shows the distribution of all the calculated distances. It shows that, as expected, the majority of all the weekday AM peak trips boardings take place near the reported address: Almost 80% of the boardings are made within a 2 km distance of the reported address. There is also an important percentage of the trips, approximately 13%, which shows a distance greater than 5 km. These large distances can be caused by an incorrect or outdated address, or by a far transfer in the second leg of a trip or even a drop-off. Ultimately, only those boardings with a distance of less than one kilometer were selected. This is equal to 50.7% of the records and trims the sample to 38,592 trips.

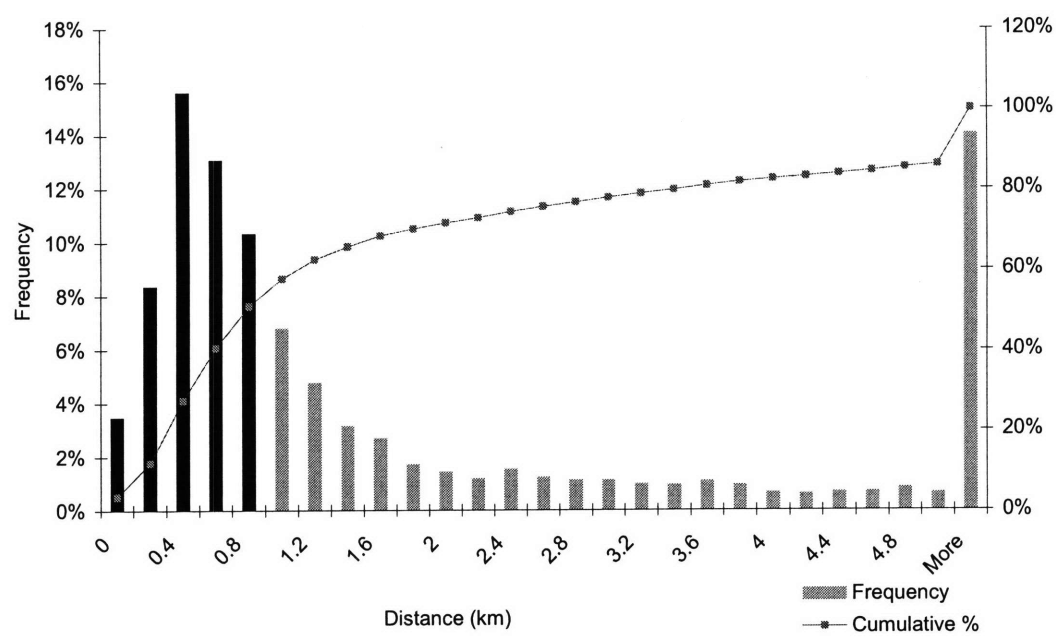


Figure 14: Distances between reported address and boarding location (N=68,938)

4.2.5 Step 6: Select first trips

One last filter of the process selects the first trips of the day. So far, the boardings have been selected based on day, time of day and distance concerns. However, it is possible to have transfers in the AM Peak and under a 1 km distance threshold as well.

Boarding #	Frequency	Percentage %
1	37738	97.79
2	817	2.11
3	24	0.07
4	8	0.02
More	5	0.01

Table 24: Frequency of multiple boardings per card

A closer examination of the trips in Table 24 shows that the majority of the selected weekday AM trips, approximately 97.8%, correspond to the first trip of the day. The rest of the trips represent either transfers or consecutive trips made by more than one person with the same card. Therefore, only the first trips (37,738) are selected to study users' boarding patterns.

4.2.6 Step 7: Select consistent users

The next step in the selection process corresponds to the aggregation of the weekly boardings per Smart Card. The purpose is to identify boarding patterns for the residents around rail stations. The current sample of 37,738 boardings represents the activity of 11,825 cards in the system With the combined AFC-AVL data it is possible to understand basic aspects of travel behavior, such as trip frequency, time consistency and modal selection.

4.2.6.1 Trip frequency

Since the data that has been selected so far corresponds to a week of boardings, it is possible to determine how many morning trips were made by user in the week. An examination of the trip frequencies per user is shown in Figure 15, where the cards are segmented according to the respective monthly dataset.

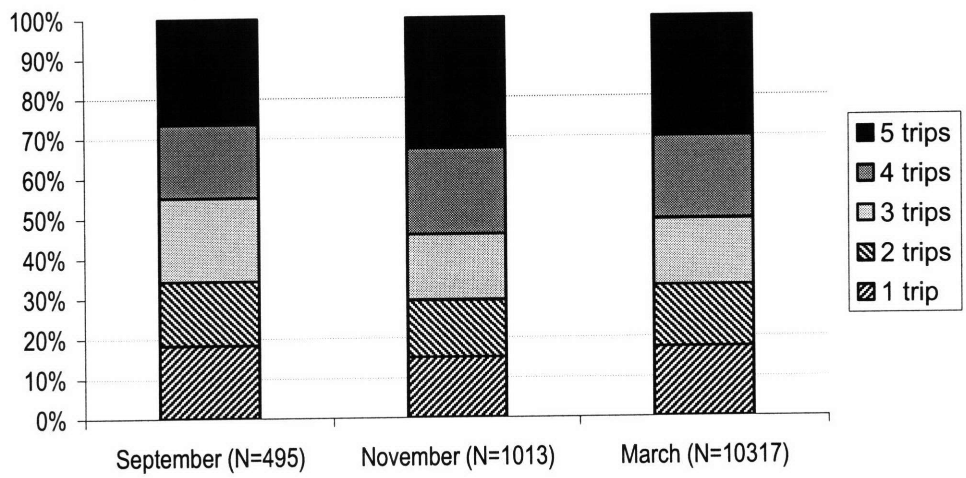


Figure 15: Weekly trips per card

As can be seen, the percentages for all months are fairly similar. Approximately 50% of the sample makes four or more trips per week. This is to be expected because the sample is taken from A.M. peak hours, when most people start their commute journeys. In order to extract a sample of day-frequent customers, only those cards used at least three times per week were selected. This selection trimmed the sample to 7,973 cards

4.2.6.2 Time consistency

In order to better identify a commuter, an additional control is established, based on the consistency of boarding times. Most commuters need to reach their daily destinations at regular hours, and because AFC data records the exact boarding date and time, it is possible to select those travelers who board the system consistently and mimic a commuter's behavior.

The criteria used for this selection was developed by Gupta (2006)[34], who defined commuters based on time and day consistency for the CTA. A time consistent user is one who regularly boards the system within a defined window of time. In this research, the window of time is restricted to 30 minutes and applies to all those customers that used the system at least three times per week as selected in section 4.2.6.1

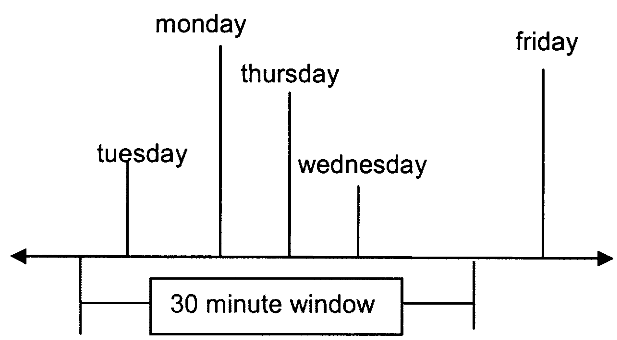


Figure 16: Sample weekly boardings of a 4 day-consistent user

A customer who boarded the system five days a week can be 5, 4 or 3 days consistent. A customer who boarded it four days in the week can be 4 or 3 days consistent and a customer who used it three days in the week can only be 3 day consistent. As a general rule, if at least three of the weekly boardings were not made within a 30 minute boundary, then the user is ruled as non-consistent. Figure 17 shows the distribution of time consistency users for each month.

[34] Gupta, Saumya (2006) Understanding Transit Travel Behavior: Value added by Smart Cards. MIT thesis

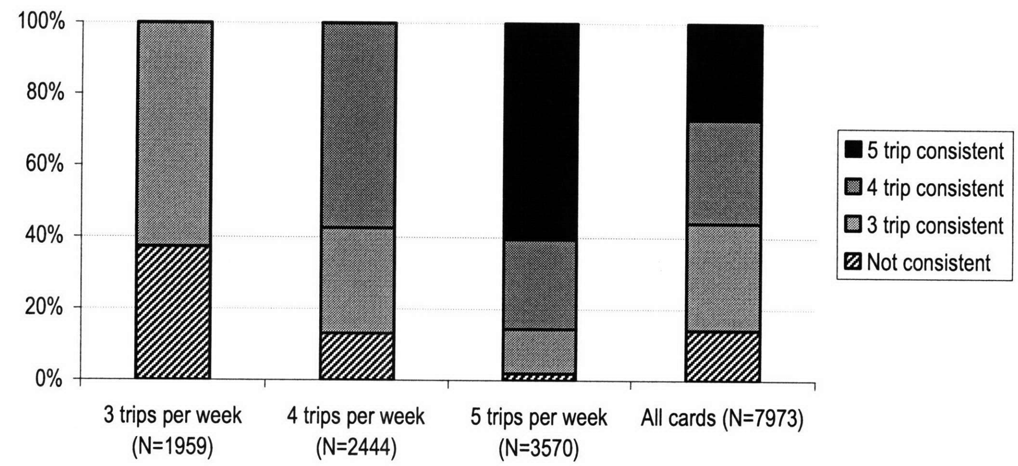


Figure 17: User trip consistency

In general terms, 86% of the users are considered consistent in some degree. This is expected due to the times at which the boardings are selected and the frequency (3 or more trips per week) of card usage. Those users who are not considered consistent might have more flexible work schedules or just different workplaces so that it allows them to have many boardings per week at an irregular time pattern. For the purposes of this research, the consistent users are considered commuters and represent the group of users who become the subject of analysis. After this filter was applied, the sample was trimmed to 6,853 cards total for the months of September, November and March.

4.2.7 Step 8: Develop user specific profiles

After having processed the data from the AFC and AVL systems it is possible to describe the boarding patterns of the selected users. As described in section 4.1.5, the variables that are being studied after the reductions in level of service, are frequency of use, modal selection, boarding time and boarding location. Therefore, a description of such travel patterns is presented below:

4.2.7.1 Trip Frequency

The majority of the selected users make 5 trips per week, accounting for up to 51% of the observations. All users were selected as time frequent users, and hence represent expected commute patterns. Figure 18 shows the distribution of trips per week across all observations

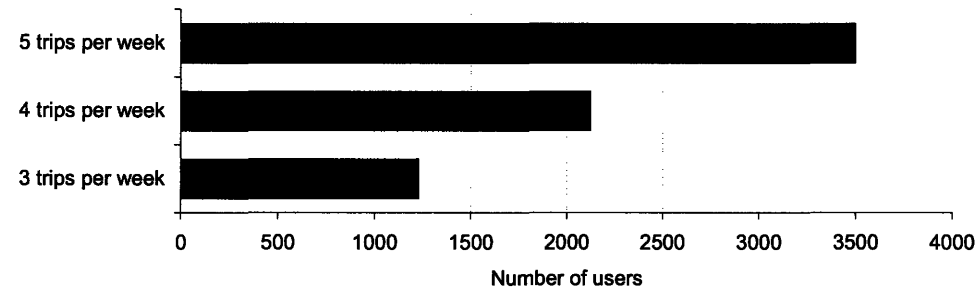


Figure 18: Trip frequency of Smart Card commuters AM peak

4.2.7.2 Modal Preference

A metric was developed to describe a user's first trip preference for train or bus. The modal preference is the percentage of weekly trips made in a particular mode. Figure 19 shows the distribution of preferences across the current selection of cards. It can be seen that most of the cards (87%) belong to rail frequent users. These, are commuters that do more than 60% of their first boardings per week into rail. The remaining cards (13%) have a stronger preference for bus, using routes that are near the affected stations. Only rail frequent users will be included for analyses of changes in travel behavior.

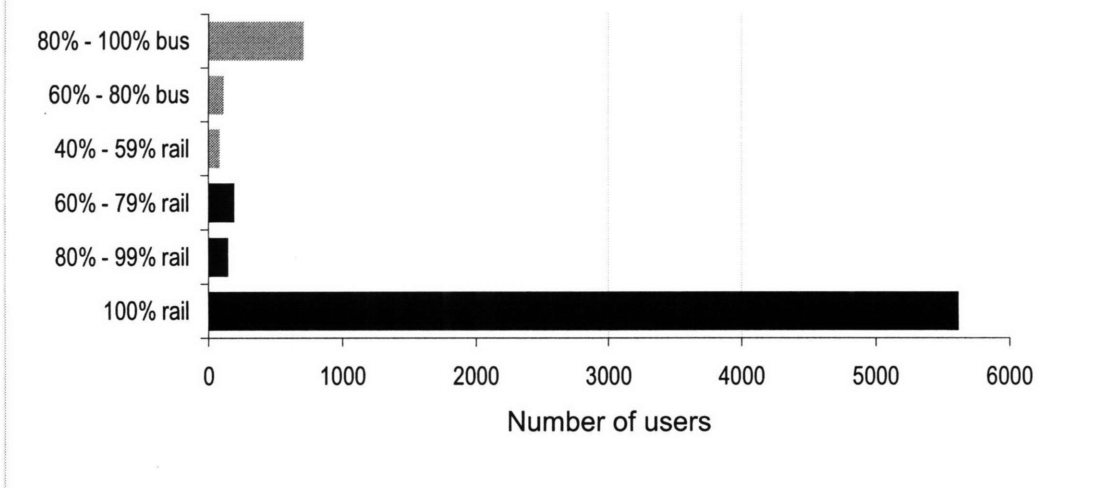


Figure 19: Modal preference of selected commuters AM peak

4.2.7.3 Boarding time

Another variable of interest for the analysis is a potential change in boarding times. Weekly boarding times were averaged for each commuter, generating an average boarding time for each card. Figure 20 shows the distribution of the average boarding times for peak hour between 6:00 A.M and 10:00 A.M. The segment between 6:30 and 9:30 captures 94% of the total passengers and the segment between 7:30 and 8:30 captures almost half of the total morning boardings.

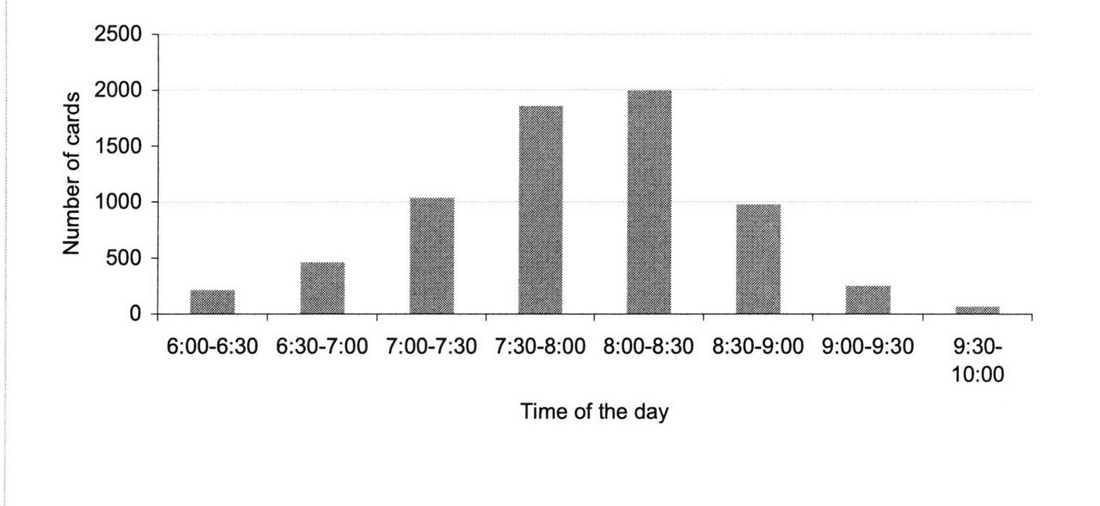


Figure 20: Boarding time distribution for commuters AM peak

4.2.7.4 Boarding Location

The boarding location is another important variable to analyze. Based on the boarding records, it is possible to assign a "home" station to each of the users. This home station represents the most frequent boarding location (or route in the case of buses) across the weekly boardings.

Rail Station	Line	Dataset	Maintenance Event	Card Users
Francisco	Brown Line	September	Station Closure	150
Kimball	Brown Line	September	Station Closure	79
Addison	Brown Line	November	Station Closure	219
Montrose	Brown Line	November	Station Closure	233
Davis	Purple Line	March	Three track operation	22
Sheridan	Red Line	March	Three track operation	358
Main	Purple Line	March	Three track operation	66
Diversey	Brown/Purple Line	March	Three track operation	426
Wilson	Red Line	March	Three track operation	142
Armitage	Brown/Purple Line	March	Three track operation	351
Linden	Purple Line	March	Three track operation	13
Argyle	Red Line	March	Three track operation	125
Wellington	Brown/Purple Line	March	Three track operation	330
Fullerton	Red/Brown/Purple Line	March	Three track operation	416
Loyola	Red Line	March	Three track operation	115
Paulina	Brown Line	March	Three track operation	391
Belmont	Red/Brown/Purple Line	March	Three track operation	650
Bryn Mawr	Red Line	March	Three track operation	190
Irving Park	Brown Line	March	Three track operation	308
Western	Brown Line	March	Three track operation	213
Other stations/bus	N/A	N/A	N/A	2056
			TOTAL	6853

Table 25: Boarding locations for selected commuters AM Peak

As shown in Table 25, most of the commuters (69.9%) have their home station in one of those stations which were affected by the maintenance project and are expected to change their behavior after the reductions in level of service. Those whose home station was not affected by BLCEP, will not be subject to further analyses of changes in travel behavior.

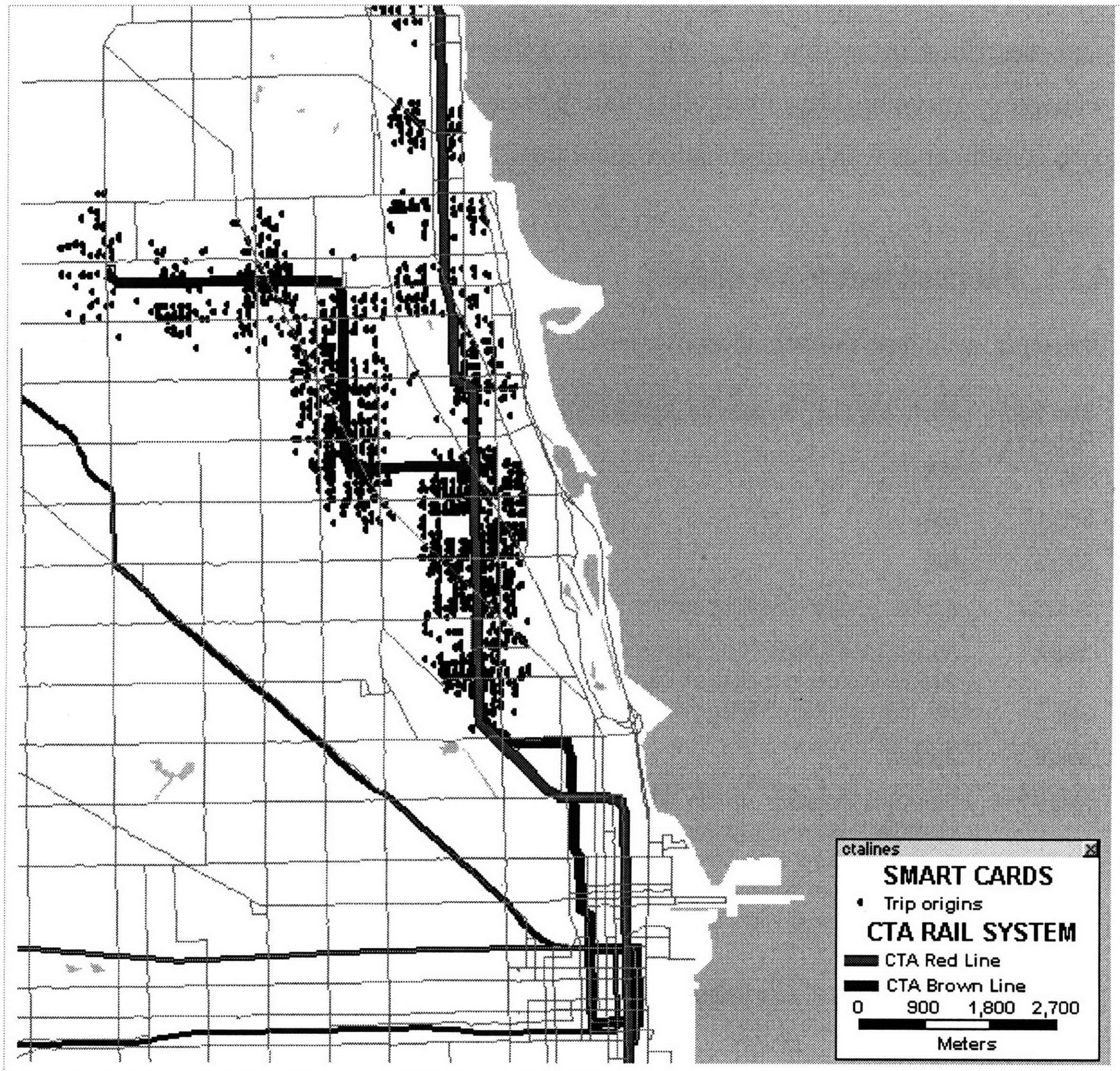


Figure 21: Spatial location of selected commuters

4.3 Other time periods

In order to examine changes in travel behavior, this research examines Smart Card activity for three additional time periods:

- The P.M. peak in the month *before* the project
- The A.M. peak in the month *during* the project
- The P.M. peak in the month *during* the project.

The procedure for extracting and analyzing the data is similar to the one used for the A.M. peak described in section 4.2. The main differences are described in the following sections. Analyzing data from other time periods provides origins and destinations for each commuter as well as information about their boarding patterns.

4.3.1 PM peak before the project

In order to calculate the behavioral variables

Step from Figure 11	Application in P.M. peak *before* the project
Step 0	N/A
Step 1	N/A
Step 2	N/A
Step 3	Applies, using the resulting card list from the AM Peak Step 7 as an input for this step.
Step 4	Applies for a different time segment. The P.M. peak is broadly defined between 3 P.M. and 10 P.M in order to capture as many returning users as possible.
Step 5	N/A (the return trip is independent from the reported home address)
Step 6	Applies
Step 7	Step 7 is modified: In the A.M. peak this step controls for the boarding patterns of a customer. However, for the afternoon period, the boarding times will presumably be too spread to control for a time window. Instead, the interest is focused on inferring the most probable return boarding location of the return trip. Section 4.3.1.1 covers in detail the procedure used to infer the location of the return trip based on the available data.
Step 8	Applies

Table 26 : Key differences in data manipulation with respect to the before period

4.3.1.1 Afternoon boarding location

One of the main limitations to replicate the travel patterns in Chicago with Smart Cards is the lack of an exit control in rail stations and bus stops. Passengers only tap in when they enter the system but there is no information on their exit location. Hence, it is impossible to know with certainty the morning trip destination.

In order to circumvent this limitation, this research uses the location of the afternoon boardings as a proxy for the destination of the morning trip. Although it is a rather narrow way to deal with the complexity of the trips, it is a reasonable approximation to explain commute trips which are commonly bounded by work hour restrictions.

Another implicit assumption is that the trip ends in the station/stop. Without further information related to the actual location of the destination, it is impossible to relax this assumption. However, Chicago Card registration forms currently do not provide this data or the address of the work center. The practical implication of this assumption for this research is that the egress walking time will be null when analyzing trip attributes.

Just as in the case of the A.M. peak trips, there is a weekly log of information for each user which allows a rich examination of the variability of the afternoon boarding location. Unlike the morning trip, where it is expected that the commuter's boarding location will not vary substantially, the afternoon trips are more likely to vary the boarding location due to constraints of their job, links with trips of other purposes, etc. Hence, in order to develop user-specific profiles, a set of rules was designed to control for the spatial variability and assign a unique return boarding location for each commuter:

- A user must have recorded at least three afternoon boardings per week. Otherwise he/she is ruled as afternoon inconsistent
- The coordinates for the *estimated* boarding location will have the average values of the X and Y coordinates of the weekly boardings. However, since there are users with 3, 4 or 5 boardings per week, there are many combinations of locations that could be used. Therefore, the rule is that the combination of boarding locations with the minimum average distance between them, will be the one that determines the *estimated* boarding location.
- The minimum average distance between boarding locations has to be at least 1.5 km, otherwise, the user is ruled as afternoon inconsistent

$$D(km) = k \times \frac{\sum^{n} \sqrt{([CC\_Xi] - [CC\_Xj])^2 + ([CC\_Yi] - [CC\_Yj])^2}}{n}$$

D(km) is the average distance between boarding locations in Kms for a given combination of n boarding locations

The selected combination is the one where:

D(km) is the minimum across all possible combinations

D(km) <= 1.5 km

The combination consists of at least 3 days (n => 3)

CC_X i is the X coordinates of the boarding location of day i

CC_Y i is the Y coordinates of the boarding location of day j

CC_X j is the X coordinates of the boarding location of day i

CC_Y j is the Y coordinates of the boarding location of day j

n is the number of boarding locations of a given combination

K = 113.325 is the constant that converts distances from XY coordinates to kilometers

Equation 4: Decision rules for estimating the afternoon boarding location

After estimating the afternoon boarding locations, the resulting coordinates were included in a map to check for a preliminary snapshot of the resulting job locations. As can be seen in Figure 22, most of the presumed jobs are in the Loop and its vicinity. This is according to expectations as this area is where most of the jobs are located and where all train lines converge. There is some concentration of trips around the shared portions of the Red-Brown Line, which is also reasonable because of the mixed use of the Lincoln Park area.

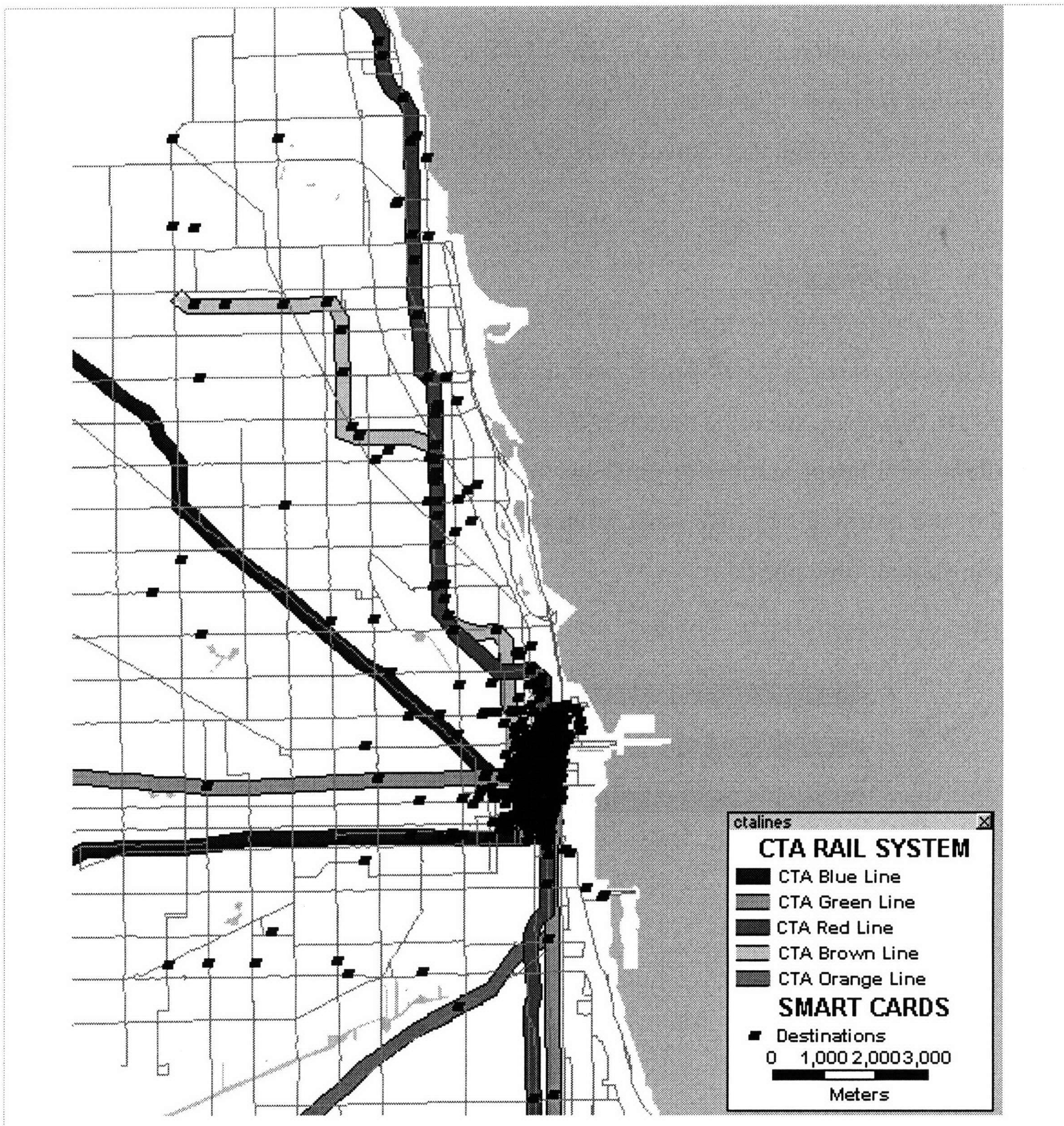


Figure 22: Estimated morning destinations

4.3.1.2 User profiles

In the same way as was done with the A.M. peak, user profiles were developed for the P.M. peak. First, this partially complements the morning trip profile by assigning a destination to it. Second, this sets a baseline to be compared with the time period *during* the project.

PM profiles were developed for 4,797 cards, which represent those AM commuters that had their home station affected by the BLCEP (See Table 25) and used rail for more than 60% of their AM trips (as seen in Figure 19). This sub-set represents those commuters that are likely to change their travel behavior due to the BLCEP.

4.3.1.2.1 Trip Frequency

Figure 23 shows how the majority of the selected commuters used the system at least three times per week in the PM peak period. This distribution is similar the one shown in Figure 18. However, it can be seen that there is a number of users who ride the system in the afternoon a lower number of times per week than in the morning. This suggests that some of the morning trips use other modes for their return leg, other fare media or take place in another time period.

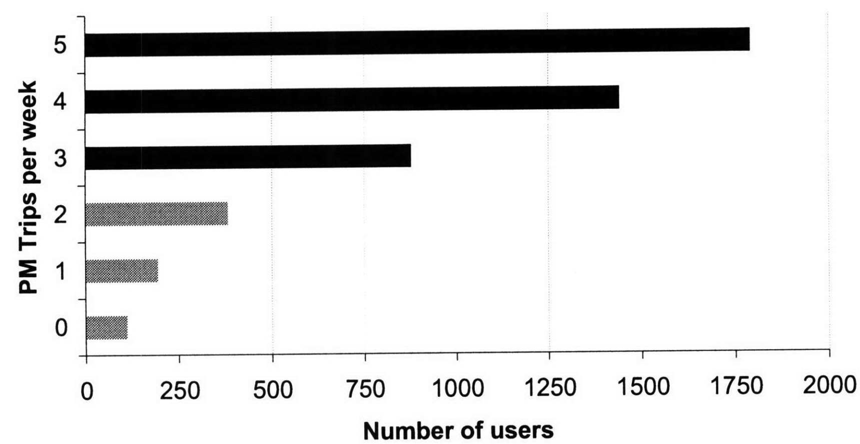


Figure 23: Frequency for commute return trips PM peak

4.3.1.2.2 Modal preference

Figure 24 shows the distribution of the modal selection across commuters for the return trip. As can be seen, the selected commuters have a strong preference to make all their return trips by rail. 79% of them use rail for 100% of their afternoon weekly boardings and 91% of them use rail for 60% or more of their afternoon weekly boardings.

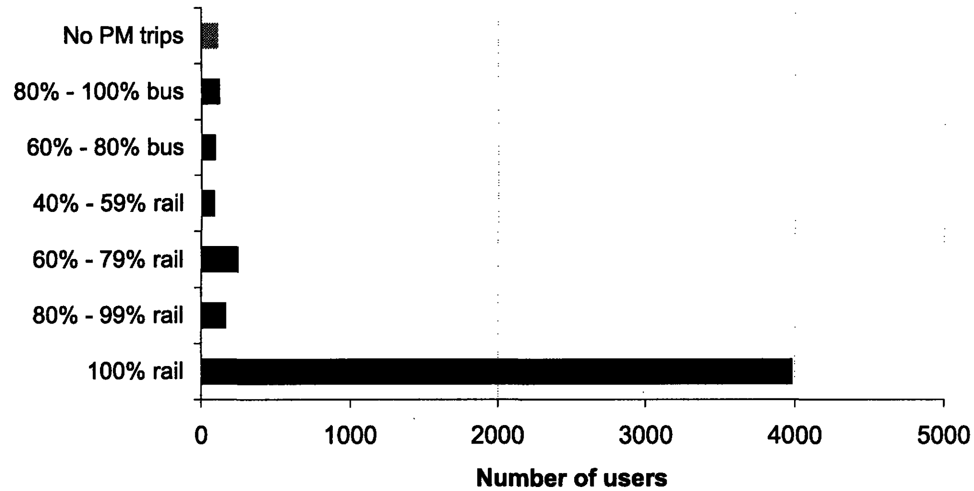


Figure 24: Modal selection of return trips PM peak

4.3.1.2.3 Boarding time

Average boarding times in the afternoon were also calculated for each card. The PM peak was initially broadly defined between 3 P.M and 10 PM. However, as it can be seen in Figure 25, the segment between 4:00 PM and 7:30 PM captures 92% of all the cards. The remaining 8% is comprised in part by boardings at other times of the afternoon (6%) and by cards that did not register boardings in the afternoon (2%). This distribution is consistent with the expected patterns for the peak hour.

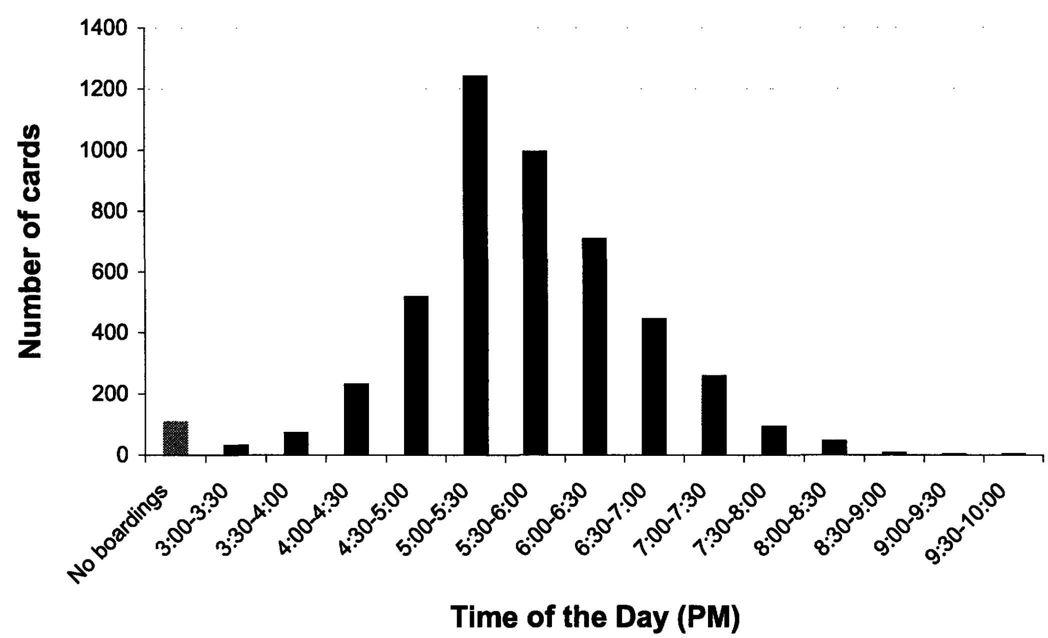


Figure 25: Average boarding times of return trips PM peak

4.3.1.2.4 Boarding location

The final parameter to examine for the PM peak is the boarding location of the return trips. Each user was assigned a station and/or a stop depending on his/her boarding habits: The boarding location that was most often used by that commuter in that week is considered his/her "home" return location. For users who are 100% rail users or 100% bus users, only one station or stop applies. For users who have mixed preferences, both a station and a stop was assigned. As seen in Figure 24, there is a marked predominance to use rail for 100% of the return trips. However, having a detailed inventory of the most used bus routes will set a baseline for comparison to see which are the routes which can serve effectively as substitutes for the affected rail lines.

Rail station	No. Users	% of all
Washington/Wells *	618	13.53%
Quincy/Wells *	592	12.97%
Merchandise Mart	394	8.63%
Monroe/State *	380	8.32%
Lake/State *	358	7.84%
Clark/Lake *	279	6.11%
Grand/State	236	5.17%
Chicago/State	204	4.47%
State/Lake *	190	4.16%
Jackson/State *	187	4.10%
LaSalle/Van Buren *	152	3.33%
Chicago/Franklin	135	2.96%
Adams/Wabash *	120	2.63%
Randolph/Wabash *	115	2.52%
Madison/Wabash *	74	1.62%
Davis	70	1.53%
Library *	51	1.12%
Fullerton	29	0.64%
Clark/Division	27	0.59%
Clinton-Congress	23	0.50%
Harrison	22	0.48%
Polk	19	0.42%
Foster	18	0.39%
North/Clybourn	16	0.35%
Clinton-Lake	15	0.33%
Armitage	14	0.31%
Other 60 stations	228	4.99%
TOTAL:	4566	100%

* Station belongs to the Loop

Table 27: Rail boarding locations. PM peak

As it can be seen in Table 27, the majority of the rail boardings for the PM peak are made in the Loop. Among the 15 most used stations, 11 belong to the Loop and the other four are no more than two stations away. This is according to expectations and is consistent with the map shown in Figure 22

Bus Route	No. Users	% of total
151	58	8.23%
147 *	49	6.95%
22	45	6.38%
36	36	5.11%
156	34	4.82%
66	32	4.54%
20/X20	28	3.97%
146 *	21	2.98%
8 / X8	18	2.55%
157	18	2.55%
11	17	2.41%
148 *	17	2.41%
145 *	15	2.13%
56	12	1.70%
125	12	1.70%
134	10	1.42%
135	10	1.42%
3	9	1.28%
14	8	1.13%
29	8	1.13%
50	8	1.13%
9	7	0.99%
65	7	0.99%
77	7	0.99%
126	7	0.99%
173	6	0.85%
2	5	0.71%
6	5	0.71%
12	5	0.71%
76	5	0.71%
80	5	0.71%
144	5	0.71%
62	4	0.57%
Other 32 routes	4	9.9%
Undetermined bus route	102	14.47%
TOTAL	705	100%

* Lake shore drive express route

Table 28: Bus routes PM peak

Table 28 shows the distribution of passengers across bus routes and is consistent with the spatial location of passengers and jobs. Comparing these routes to the system map it can be seen that the ten most used bus routes are a connection to the Loop or to its vicinity.

Eight of those ten routes are direct North-South connection and two of them (20/X20 and 66) are East-West, probably serving as connectors to rail stations. It is also noticeable the high percentage of bus boardings whose bus route was impossible to determine. This is low match rates between bus numbers from the AFC and AVL records, as explained in section 4.1.4. However, this analysis is valid assuming that the matching rate is equal for all bus routes. Further research can be made to identify the validity of this assumption.

4.3.2 AM and PM peak *during* the project

After having developed the complete user profiles for the time period *before* the project, a very similar procedure is applied for a time period *during* the project. The idea is to evaluate the same behavioral variables for the same users under a worsened rail level of service.

The BLCEP presents an opportunity to evaluate two different types of changes. The first is closing stations and tracks under repair, reducing average train speeds and increasing walking times. Another effect that can be evaluated is the return to normal conditions, once the stations are reopened. The weeks selected to evaluate such changes are spaced between one and four weeks after the service changes as shown in Table 29.

Type of change in service	Date of event	Effects in the travel experience	Week to evaluate behavioral changes
Station is closed	Kimball and Francisco (09/15/06)	Longer walking times	October 16th – 20th / 2006 Four weeks after closure
Station is closed	Montrose and Addison (12/02/06)	Longer walking times	December 4th – 8th / 2006 One week after closure
Three track operation	All stations in the Brown Line and Red Line (04/02/07)	Longer waiting times, Longer travel times	May 1st – 7th / 2007 Four weeks after three track started
Station is reopened	Kimball (01/12/07)	Walking times return to status quo	February 8th – 14th / 2007 Four weeks after reopening
Station is reopened	Francisco (03/09/07)	Walking times return to status quo	March 17th – 23rd / 2007 One week after reopening
Station is reopened	Montrose (11/26/07) and Addison (12/03/07)	Walking times return to status quo	December 10th – 16th / 2007 Between one and three weeks after reopening

Table 29: The Brown Line Capacity Expansion Project. September 2006-June 2007

In terms of the data extraction and processing, the methodology framework is explained in section 4.2 and illustrated in Figure 11. However, Table 30 shows key differences worth mentioning .

Step from Figure 11	Application in A.M. peak *during* the project	Application in P.M. peak *during* the project
Step 0	Download and link the AFC and AVL data of the corresponding month and week	Download and link the AFC and AVL data of the corresponding month and week
Step 1	Does not apply	Does not apply
Step 2	Use list of cards that resulted from the AM peak *before* the project	Use list of cards that resulted from the AM peak *before* the project
Step 3	Applies with the abovementioned list of cards	Applies with the abovementioned list of cards
Step 4	Applies	Applies between 3 p.m. and 10 p.m.
Step 5	Does not apply.	Does not apply
Step 6	Applies	Applies
Step 7	Aggregate boardings to users but do not restrict the list to inconsistent users	Aggregate boardings to users but do not restrict the list to inconsistent users. In the same way as in the P.M peak period *before* the project, check that the estimations of the afternoon boarding location are reliable
Step 8	Applies	Applies

Table 30: Key differences in data manipulation with respect to the *before* period

The results of processing these data are the new travel profiles for the same users that were selected in sections 4.2 (AM peak *before)* and 4.3.1 (PM peak *before).* The completion of the data for the *during* period, facilitates an individual examination of the variables of interest, which in turn, serve to explain changes in travel behavior. The results of examining these changes are discussed in Chapter 5

4.3.3 Control zones

One of the chief limitations of using smart cards to analyze behavioral changes over time is the reliance on the existence of a physical card. As mentioned, one of the basic assumptions in this research is that a customer is uniquely represented by a card in the system. However, if this card is misplaced, damaged or transferred to another individual, then the analysis assumed for one person will become incomplete or biased.

In order to overcome this limitation, this research takes as an assumption that there is a natural occurrence of card losses and replacements in the system. This means that every day a certain number of cards will be taken off of circulation due to different reasons and that the user will either replace it or change his fare media. In practical terms, this means that some of the passengers that will not report any activity in the period *during* the project will still be traveling but will not be monitored.

One way to quantify this process is by through the analysis of the rate of card losses across commuters in a control zone. The ideal control zone is one where the land use characteristics, the transportation level of service and the socio economic conditions are similar to the zone under study. However, in a city like Chicago it is very difficult to find these equalizing conditions. Instead, this research analyzes how this dynamic played out for the commuters that live around the stations of the Orange, Green and Blue lines. A 1 km buffer was created around these lines; these buffers will be called from now on the Control Zone, as depicted in Figure 26.

The analysis of the control zone is similar to the one applied to the stations of interest. It requires the development of user specific profiles for the AM and PM peak *before* and *during* the BLCEP, under the same conditions as explained in sections 4.2, 4.3.1 and 4.3.2. We will control for the percentage of cards that stop reporting activity and the time elapsed between the two time periods. This gives us a baseline to be compared against the number of cards that did not report any activity in the areas where station closures and degraded track operations occur.

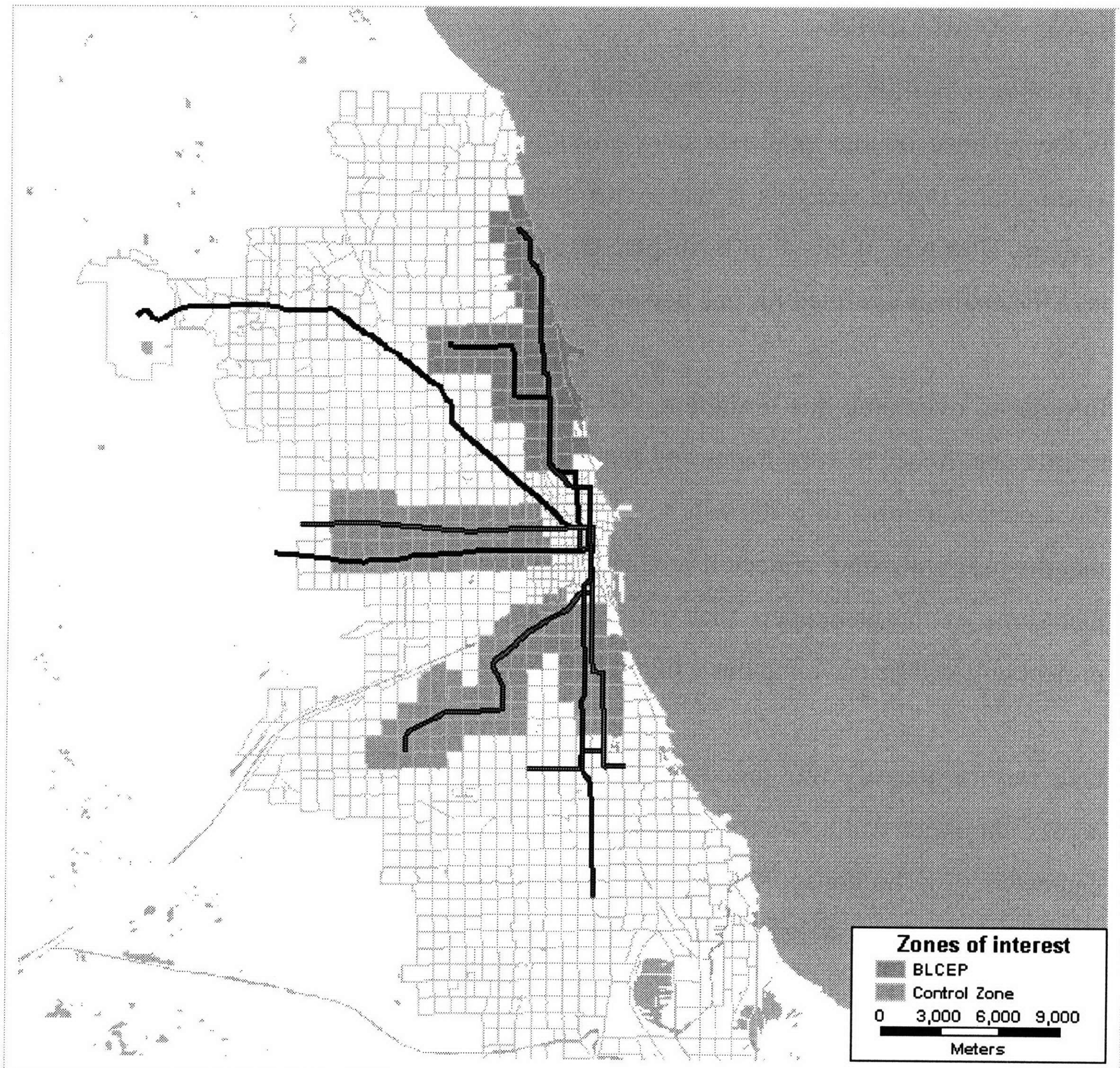


Figure 26: Zones of interest

As shown in Table 31, the stations in the control zone provide a reasonable number of observations to examine the dynamic of card losses. Although the number of commuters per station is lower, this is compensated by a higher number of analyzed stations. This finding is consistent with the lower use of smart cards in the south and west sides of Chicago, as shown previously in Figure 9.

Table 31 also shows that the patterns of commuting are fairly similar across the study zone and the control zone. The boarding times and trip frequencies are comparable across zones and in both cases show a higher use of public transportation for the morning trips rather than for the return trips. Boarding times are similar for the morning trips but slightly later in the afternoon for the study zone.

Commuters in area of maintenance project

			AM peak		PM peak	
	N	No. Stations	Average trip frequency *	Average boarding time	Average trip frequency *	Average boarding time
September	229	2	4.28	7:58 A.M.	4.11	5:29 P.M.
November	452	2	4.49	7:52 A.M.	4.14	5:33 P.M.
March	4116	16	4.47	7:58 A.M.	4.10	5:44 p.m.

Commuters in control zone

			AM peak		PM peak	
	N	No. Stations	Average trip frequency *	Average boarding time	Average trip frequency *	Average boarding time
September	1120	28	4.46	7:54 A.M.	4.27	5:18 P.M.
November	1096	28	4.41	7:57 A.M.	4.11	5:20 P.M.
March	1810	28	4.62	7:56 A.M.	4.25	5:19 P.M.

* Trips per week

Table 31: Descriptive statistics control and study zone

4.4 Limitations of the presented approach

This chapter has shown how we selected a base of customers to study different aspects of travel behavior. However, as shown, the filtering process separated a fair amount of records that represent other customers as well.

The advantage of using the filtering to select commuters-only is that a group of commuters is likely to repeat the same trip *before* and *during* the maintenance project. Therefore, if he/she makes the same OD pair in both time periods, it is likely that some changes in travel behavior can be attributed to BLCEP. However, In the case of non commute trips, it is more complicated to assess the destination of a trip (because there is no expected return time, like the PM peak) and it is difficult to find the same discretionary trip in two different time periods.

The downside of selecting only commuters is that the results can not be generalized to all the base of CTA customers. First, the time spent in work trips tends to be highly valued compared to non-work trips so the changes in behavior may be different as well. Also, the crowding conditions of the peak hours may influence the behavioral responses. Therefore, the results of these research should be seen under the light of the analysis of the commuter market as opposed all the CTA users.

5 Changes observed in travel behavior

This chapter draws on the description of the commuter's behavior *before* and *during* the maintenance project as developed in Chapter 4. A comparison of both instances is made to reveal what type of changes the project induced in the passengers' travel behavior. Changes in the following variables are examined based on the previously developed user profiles:

- Modal selection.
- Trip frequency.
- Boarding locations.
- Boarding times.

These comparisons will hinge on the assumptions that a smart card is generally owned by an individual and is not transferred over time to another person. Although it is impossible to determine whether this is true or not for each card, some controls are discussed in section 6.4 to enhance the validity of this assumption.

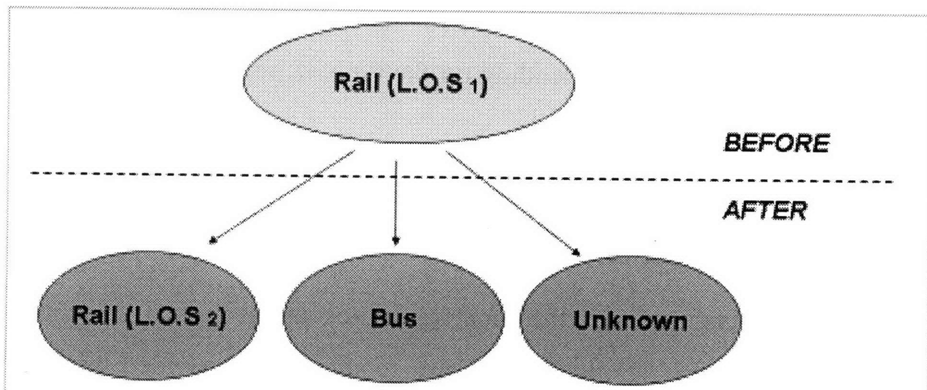
As summarized in Figure 2, the Brown Line Capacity Expansion Project is a staged project, which in practical terms means that not all stations are closed at the same time. The same applies for three track operation, which does not occur simultaneously with the station closures. Hence, the described comparisons made in this chapter have different *before* and *during* in-between times depending on the type of impact and location that is analyzed.

5.1 Changes in modal selection

Probably the behavioral change with the most relevance is the modal selection. The agency must decide how much supplementary service should be provided based on the expectations of passengers using the bus, due to the inconvenience caused by the rail service. On the other hand, there are financial concerns due to drop in revenue caused by passengers who leave the CTA, and, to a lesser extent, from those who use the bus, as they pay lower nominal fares.

In order to quantify these modal shifts as accurately as possible, it is important to clarify the extent and characteristics of this analysis:

- All the passengers in the period *before* were rail commuters. This means that they used rail in the AM peak hour on a regular weekly basis.
- All the passengers lived in the vicinity of the affected stations before the maintenance project started.
- The modal decision *during* the project involves three choices, as shown in Figure 27: A. Those who continued being frequent rail users despite an imposed penalty in their trip experience. B. Those who switched to the bus as frequent users and, C. Those whose cards did not register any activity and that include people that used other modes or other fare media, lost their cards, lost their jobs, etc.
- The changes in modal selection *during* the project can only be represented under different levels of certainty and aggregation. For groups A and B, we have 100% certainty about the mode selection for each individual. However for group C, since there is no actual indication of what they did, we can only establish if the share of cards that did not register activity is larger than the share in the control zone. This could suggest, in aggregate terms, that there was a shift towards using other non-transit modes.



Where L.O.S. 1 > L.O.S 2
Figure 27: Observed decisions from examining smart cards.

Station closures[35]

Station	Line	N*	RAIL	BUS	UNK**	RAIL%	BUS%	UNK%
Kimball	Brown Line	79	69	4	6	87.3%	5.1%	7.6%
Francisco	Brown Line	150	125	14	11	83.3%	9.3%	7.3%
Addison***	Brown Line	219	178	31	10	81.3%	14.2%	4.6%
Montrose***	Brown Line	233	182	29	22	78.1%	12.4%	9.4%
Total station closures		681	554	78	49	81.4%	11.5%	7.2%

Three track operation

Station	Line	N*	RAIL	BUS	UNK**	RAIL%	BUS%	UNK%
Davis	Purple Line	22	13	1	8	59.1%	4.5%	36.4%
Sheridan	Red Line	358	284	40	34	79.3%	11.2%	9.5%
Main	Purple Line	66	49	3	14	74.2%	4.5%	21.2%
Diversey	Brown/Purple Line	426	348	31	47	81.7%	7.3%	11.0%
Wilson	Red Line	142	108	11	23	76.1%	7.7%	16.2%
Armitage	Brown/Purple Line	351	289	19	43	82.3%	5.4%	12.3%
Linden	Purple Line	13	12	0	1	92.3%	0.0%	7.7%
Argyle	Red Line	125	94	11	20	75.2%	8.8%	16.0%
Wellington	Brown/Purple Line	330	274	21	35	83.0%	6.4%	10.6%
Fullerton	Red/Brown/Purple Line	416	345	22	49	82.9%	5.3%	11.8%
Loyola	Red Line	115	90	9	16	78.3%	7.8%	13.9%
Paulina	Brown Line	391	293	34	64	74.9%	8.7%	16.4%
Belmont	Red/Brown/Purple Line	650	519	53	78	79.8%	8.2%	12.0%
Bryn Mawr	Red Line	190	152	20	18	80.0%	10.5%	9.5%
Irving Park	Brown Line	308	243	23	42	78.9%	7.5%	13.6%
Western	Brown Line	213	173	9	31	85.4%	4.2%	14.6%
Total three track operation		3901	3071	307	523	78.7%	7.9%	13.4%

* N is the number of commuters that were examined per station after the filtering process
** UNK: A passenger is considered as UNK when his card does not report activity in the period *during*
*** The *during* period for these two stations is two weeks after the closure. For all other stations is four weeks

Table 32: Modal share after station closures and three track operation

Table 32 shows the extent of the modal switch after station closures and during three track operation. For the purposes of classification, passengers are segmented as rail users, bus users or unknowns. A rail user is one who made at least half of his/her first trips in rail, while a bus user is one who did more than half of his/her trips in bus. Unknowns (UNK) are those passengers without travel records in the period in which the construction projects were active.

[35] These 4 stations were closed, therefore, those passengers reported as 'rail' had to use another station in the period during construction.

As can be observed, the majority of the passengers continued to use rail. The average rate of continuing (rail) passengers is almost the same for three track operation as for the station closures. The average for all the examined stations is 79.1% (3625 out of 4582 passengers). Only two stations had a rate of rail users below 60% (Davis, Wellington) and one of these has a rather small sample (Davis N=22). These two stations also had the highest rates of Unknown users (30%-36%), suggesting that the real rate of rail continuing passengers could be closer to the average.

On the other hand, there is a noticeable increase in the use of the bus by commuters. The share of users who switched to bus varies between 0% and 18%, depending on the station. Stations with a better bus service will likely witness more passengers switching to bus. In section 6 an attempt will be made to explain mode shift. The project average was 8.4% and, with the exception of the station Linden (N=13), all stations saw at least 4.5% of their passengers make a mode switch.

Finally, the rate of Unknown users is rather variable across stations. It can be as high as 30.4% in the Wellington station or as low as 4.6% in Addison. However, there are major differences that have to be controlled for.

- First, as mentioned in a previous section, there is a natural turnaround of cards that are lost or damaged due to random reasons: The number of lost cards tends to increase with time, hence, for a sample of cards in time i, the number of lost cards will be grater for time $i+2$ as compared to time $i+1$.
- Second, besides the random card losses, there can be a related seasonal effect with the number of passengers who stop using public transportation. For instance, December can be a month with less recorded trips by comparison to a sample from a previous month, due to the propensity of people to leave for holidays.

The practical implication of the abovementioned considerations is that, when comparing the rate of Unknowns of the Zone of interest with the rate of Unknowns in the Control Zone, the comparison must be made for the exact same *before* and *during* time periods.

Station closures

	Time *before* and *during*			Stations		Control zone		Difference	
				N		N			
Station	*Before*	*During*	Weeks	station	%UNK	control	%UNK	Dif	t-stat
Kimball	Sept	Oct	4	79	7.59%	1120	7.32%	0.27%	0.09
Francisco	Sept	Oct	4	150	7.33%	1120	7.32%	0.01%	0.01
Sub-total Sept-Oct stations			***4***	***229***	***7.42%***	***1038***	***7.32%***	***0.10%***	***0.05***
Addison	Nov	Dec	2	219	4.57%	1096	7.12%	-2.55%	-1.37
Montrose	Nov	Dec	2	233	9.44%	1096	7.12%	2.33%	1.21
Sub-total Nov-Dec stations			***2***	***452***	***7.08%***	***1018***	***7.12%***	***-0.04%***	***0.02***

Three track operation

	Time *before* and *during*			Stations		Control zone		Difference	
				N		N			
Station	*Before*	*During*	Weeks	station	%UNK	control	%UNK	Dif	t-stat
Davis	April	May	4	22	36.36%	905	8.84%	27.52%	4.33**
Sheridan	April	May	4	358	9.50%	905	8.84%	0.66%	0.36
Main	April	May	4	66	21.21%	905	8.84%	12.37%	3.26**
Diversey	April	May	4	426	11.03%	905	8.84%	2.19%	1.25
Wilson	April	May	4	142	16.20%	905	8.84%	7.36%	2.71**
Armitage	April	May	4	351	12.25%	905	8.84%	3.41%	1.80*
Linden	April	May	4	13	7.69%	905	8.84%	-1.15%	-0.14
Argyle	April	May	4	125	16.00%	905	8.84%	7.16%	2.51**
Wellington	April	May	4	330	10.61%	905	8.84%	1.77%	0.93
Fullerton	April	May	4	416	11.78%	905	8.84%	2.94%	1.64*
Loyola	April	May	4	115	13.91%	905	8.84%	5.07%	1.74*
Paulina	April	May	4	391	16.37%	905	8.84%	7.53%	3.88**
Belmont	April	May	4	650	12.00%	905	8.84%	3.16%	1.99**
Bryn Mawr	April	May	4	190	9.47%	905	8.84%	0.63%	0.28
Irving Park	April	May	4	308	13.64%	905	8.84%	4.80%	2.38**
Western	April	May	4	213	14.55%	905	8.84%	5.71%	2.48**
Sub-total April-May stations			***4***	***4116***	***12.71%***	***905***	***8.84%***	***-3.87%***	***-3.23******

* Significant at a 90% level of confidence
** Significant at a 95% level of confidence

Table 33: Comparison of cards with unknown activity *during* BLCEP in stations and control zone

As it can be seen in Table 33, the rate of Unknowns in the control zone is between 7% and 8% after four weeks. In almost all cases it is lower than the rate of Unknowns for all stations, with the exception of Linden (N=13) and Addison. This would support the hypothesis that some commuters will ride the system less than before due to the maintenance project.

However, the picture is mixed when examining the difference between the unknown rates of a station with the unknown rates of the control zone. Using differences of proportions tests, it can be seen that this difference is not statistically significant for any of the cases when there was a station closure. In aggregate terms, when examining the cases for station closures, there is no statistical evidence to prove that smart card commuters left the system.

For the case of three track operation, there is more evidence to show that some commuters left the transit system. In aggregate terms, there is a difference of 3.87% of Unknowns with respect to the control zone. This difference is significant and gives a good idea about the extent of the passenger losses. When examining the individual stations, 11 out of the 16 stations show statistically significant differences, which range from 2.94% to 7.53% (setting aside those stations with an N<100).

5.2 Changes in trip frequency

Another variable of interest is the change in trip frequency. The previous section described how people changed their mode by defining a commuter as a rail or bus user. However, this description does not explain the impact of the project in the number of trips that a commuter makes per week by public transportation. A case can be made that a maintenance project could influence a commuter to use public transportation only on some days of the week and use other modes for the remaining. Depending on the commuter's particular habits and on the magnitude of the project's impact, he might have a higher/lower propensity to use transit only for some days of the week

Because data are available for a week of activity in two different time periods (*before* and *during*), average trip rates per week were computed for passengers and aggregated at the station level. Only passengers that reported activity with through smart card were included in this analysis, as there is no certainty about the other passengers. Table 34 shows that for all the stations -except Linden (N=12)-, there is a noticeable decrease in trips in the two periods examined regardless of the type of event. The impact for all observed commuters, shows a change from 4.40 trips per week to 3.96 trips per week, which represents a net 10% reduction in weekly boardings.

Location			Trips per week	
Station closures 2006/09	Weeks*	N**	Before	During
Kimball	4	73	4.44	3.90
Francisco	4	139	4.19	3.90
Station Closures 2006/11				
Addison	2	209	4.37	4.01
Montrose	2	211	4.49	4.08
Three track operation 2007/04				
Davis	4	14	4.36	4.21
Sheridan	4	324	4.40	3.87
Main	4	52	4.38	3.92
Diversey	4	379	4.42	3.97
Wilson	4	119	4.37	4.08
Armitage	4	308	4.34	3.93
Linden	4	12	4.23	4.25
Argyle	4	105	4.30	3.86
Wellington	4	295	4.40	4.04
Fullerton	4	368	4.50	3.93
Loyola	4	99	4.21	3.82
Paulina	4	327	4.42	3.83
Belmont	4	572	4.47	3.98
Bryn Mawr	4	172	4.35	3.98
Irving Park	4	266	4.35	3.97
Western	4	182	4.43	4.03

* Number of weeks between the *before* and the *during* periods

** Number of examined commuters

Table 34: Changes in trip frequency (commute trips per week)

However, this 10% should be examined more closely because it can not be entirely attributed to the maintenance project. Just like with the case of the Unknown passengers in section 5.1, it is necessary to compare how the trip frequency changed in other areas of the city such as in the Control Zones.

There are, however, reasons to believe that this reduction in trip frequency could also be happening in other places.

- First, the selected passengers are all people that had a frequent habit of using public transportation. This lead us to believe that they have a job and that their trips represent commute trips. However, a change in their job schedule or the loss of the job could be reflected in the trip frequency rate.
- Second, there can be cases of cards being lost in the middle of the week corresponds to the *during* phase of the project. If this is the case, then a commuter may report trips only until the day it was lost but might continue traveling in the system. This is particularly relevant given that the rate of card losses found in section 5.1 is quite high (between 7 and 8 percent for a 4 week period).
- Third, there may be seasonality and weather related effects associated with the level of economic activity that can affect a commuter's trip frequency. For instance, if in a week there was a day of bad weather then some commuters may decide to use other modes rather than walking to take transit.

Assuming that these three effects could affect different areas of the city at the same rate, then a comparison of trip frequency is made here with respect to the control zone.

Location		trips per week			
	N	Average trips made *before*	Average trip reduction *during*	Std Dev	Std Dev mean
Three track					
2007/04	3593	4.41	0.47	1.31	0.02
Control zone	825	4.45	0.36	1.17	0.04
			Difference = **0.11***		
Station closures					
2006/09	212	4.27	0.36	1.49	0.10
Control zone	1038	4.44	0.28	1.14	0.04
			Difference = **0.08**		
Station closure					
2006/11	420	4.43	0.42	0.75	0.04
Control zone	1018	4.40	0.31	1.20	0.04
			Difference = **0.11***		

* Difference is significant at a 95% level of confidence

Table 35: Difference of trip frequencies compared to control zone

Table 35 shows how changes in the trip generation rate changed for passengers affected by the station closures (September and November) and for those affected by three track operations (April). In average, all passengers did about 0.43 trips less per week. However, when comparing this trip reduction with the records in the control zone, this number becomes less relevant. Table 35 shows that, according to our expectations, recorded trips in the control zones also decreased in the period referred as *during* by 0.32 less trips per week.

This finding opens the question of whether the 0.43 and the 0.32 figures are statistically different. When performing difference of means tests for the three separate cases, it was found significant for two of them, as seen in Table 35. The difference ranges between 0.08 and 0.11 trips per week and would suggest that, due to the maintenance project, commuters used the CTA On average 2% less times per week.

This number suggests that the boarding losses caused by reduced trip frequency are small -but not fully negligible- compared to those caused by full mode shifts. For instance, if for every 1000 customers, 3.5% of them left the CTA and 97% of them continued using the system, but decreased their weekly trip frequency by 2%, then we would have an expected 942 daily boardings. Among the missing 54 boardings, 35 would be people who completely shifted to other modes and the remaining 19 would be due to a generalized reduction in trip frequency. The associated impact in revenue would even be greater when taking into account the fare differential between rail and bus. Further explanation of these scenarios is contained in Chapter 6.6

5.3 Changes in boarding times

A maintenance project will likely have effects on the boarding times of some passengers. Since commuters are usually working under a specific job schedule, a change in their trip experience could force them to adjust their boarding time in order to arrive in time to their the job.

In the case of a station closure, if a commuter decides to continue using rail by a longer walk to the nearest station, his boarding time will change depending on where the station is located with respect to the original 'home' station. If it is farther back in the line, his new boarding time should be earlier to make up for the additional in-vehicle travel time. On the other hand, if the station is closer in the line to his final destination, the new boarding time will likely be later.

In the case of three track operation, longer trip times are likely to force commuters to board their trains earlier in order to make up for the additional delay. It is also possible that the additional induced congestion in trains and stations might drive a commuter to avoid the peak of the peak and board the system even earlier.

The abovementioned behavioral changes hinge on the assumption that a commuter has efficiently budgeted his time in the period *before* and *during* the maintenance project. In the case that he had a buffer of time to plan for unexpected events, he may not change his behavior. If for instance, his job schedule is somewhat flexible, it is possible that no boarding time change will be perceived.

Zone of interest*

Before boarding Time		N	Average change in boarding time	Std_Dev	t-stat**
600	630	116	14.6	35.8	4.4
630	700	216	8.2	25.4	4.7
700	730	486	4.1	20.9	4.3
730	800	948	1.3	19.3	2.1
800	830	1114	-1.6	19.6	-2.7
830	900	539	-4.6	20.2	-5.3
900	930	141	-13.6	29.2	-5.5
930	1000	33	-29.9	44.0	-3.9

Control zone

Before boarding Time		N	Average change in boarding time	Std_Dev	t-stat**
600	630	45	26.9	59.2	3.0
630	700	50	7.2	21.3	2.4
700	730	118	3.5	15.7	2.4
730	800	208	2.5	19.5	1.9
800	830	215	-2.7	20.5	-2.0
830	900	127	-5.3	24.2	-2.4
900	930	52	-16.2	27.3	-4.3
930	1000	10	-4.1	6.4	-2.0

* Does not include station closures
** t-stat is computed to test if the average is different than zero
Table 36: Change in boarding times in zone of interest and control zone

In order to analyze changes in boarding times, it is necessary to segment the users according to their boarding times. Aggregating them all in one group can dilute some of the expected results as customers who may decide to board earlier will be mixed with others who decide to board later. Table 36 shows that the change in boarding times for time segments of 30 minutes during the peak hour. As we could expect, some time segments have more observations than others, due to the boarding times distribution shown in Figure 25. The column 'average change in boarding time' represents our variable of study. However, it has a somewhat unexpected pattern: On general, passengers who usually board the system before 8 A.M. tended to board the system later and passengers who boarded the system after 8 A.M. recorded earlier boardings.

This result is unexpected, as it is a pattern that does not correspond to our hypothesis. However, when the changes in boarding times are examined in the control zone, similar results are found: A tendency to board the system closer to the peak of the peak. In order to test if the behavior in the zone of interest is different than the one in the control zone, successive statistical tests of the difference of means are computed for all the time segments. As shown in Table 37, this difference is only significant at very low confidence levels for all boarding time segments except for the 9:30 – 10:00 AM. This suggests that there is not a strong evidence to support that the maintenance project affected boarding times. This could be linked to the somewhat weak assumption that passengers do not have an extra budgeted time and that job schedules are not flexible.

			Change in boarding times		
Original boarding time segment		N	Zone of interest	Control zone	t-stat
600	630	116	14.65	26.88	-1.30
630	700	216	8.16	7.19	0.28
700	730	486	4.10	3.46	0.37
730	800	948	1.33	2.53	-0.81
800	830	1114	-1.59	-2.75	0.76
830	900	539	-4.61	-5.27	0.29
900	930	141	-13.58	-16.25	0.59
930	1000	33	-29.91	-4.13	-3.26

Table 37: Difference of means test for changes in boarding times

5.4 Changes in boarding locations

Another aspect of interest is the change in boarding locations when a maintenance event occurs. In the case of a station closure, a change is expected to happen, either in the form of the person walking to the nearest rail station or to a suitable bus stop. In the case of three track operation, if there is a change in boarding location it is likely to be in the form of a mode shift.

Continued using rail

	Original Home Station				
New 'home' station	Kimball	Francisco	Addison	Montrose	TOTAL
Kedzie	44	80	-	-	124
Rockwell	20	38	-	-	58
Irving Park	-	-	46	-	46
Paulina	-	-	116	-	116
Damen	-	-	-	91	91
Irving Park	-	-	-	73	73
Wilson	-	-	-	7	7
Others	5	7	13	11	36
Sub-total	*69*	*125*	*175*	*182*	551

Switched to bus

	Original Home Station				
New 'home' bus route	Kimball	Francisco	Addison	Montrose	TOTAL
Total	1	10	13	7	31
93	1	-	-	-	1
82	-	3	-	-	3
78	-	7	-	-	7
152	-	-	13	-	13
148	-	-	-	6	6
78	-	-	-	1	1
Sub-total	2	*20*	*26*	*14*	62

Table 38: Changes in boarding location during station closures

Table 38 shows the selected stations and routes by commuters who continued using the system after their home station was closed. From those commuters that continued using the rail, the majority of them chose to walk to the nearest station. As it can be seen in Table 38, the stations that are listed in the first column correspond to the neighboring stations of those that were closed. However, the proportion of passengers that go to one or the other, station is not evenly distributed. It is likely to be dictated by the distribution of passengers in the catchment area, and the location of the station in the line with respect to the ultimate trip destination of each.

Table 38 also presents the bus routes that were selected by customers that shifted mode after the station closure. A closer examination of these routes shows that they are either a North-South direct connection to the Loop (Rt 148), a North-South connection to other rail lines (Rt 82) or East-West connections to other rail lines (Rt 78, Rt 152). Due to the rather small sample it is difficult to make more general conclusion at this point. However, section 6 will attempt to explain the choices through a discrete choice modeling framework.

Continued using rail

Boarding location	Number of rail users	% of rail riders
Same station	3013	93.1%
Other station	223	6.9%
Sub total rail users	3236	100%

Switched to bus

Bus route	Number of new bus users	% of all new bus users
11	46	15.0%
22	32	10.4%
135	22	7.2%
136	19	6.2%
147	16	5.2%
148	15	4.9%
8	12	3.9%
80	11	3.6%
134	11	3.6%
152	11	3.6%
151	10	3.3%
9	8	2.6%
156	7	2.3%
77	6	2.0%
144	6	2.0%
Unknown	18	5.9%
Other routes	56	18.2%
Sub total bus	306	100%

Table 39: Changes in boarding location during three track operation

Table 39 presents the changes in boarding location of all passengers affected by three track operation. On the one hand, we can be se that most (93.1%) passengers who decided to continue using rail, kept using their same station for their commute trips. The remaining 6.9% used another station which could be associated with a change in job location or change in residential location.

When examining the passengers who decided to switch to bus, we find a rather even distribution across routes. The 306 passengers who switched, were scattered into 44 different routes; the route that captured most new bus users got 15% of them. The six routes that captured most passengers were (Rt 11, Rt 22, Rt 135, Rt 136, Rt 147 and Rt 148) all being aligned along a direct North-South connection to the Loop.

5.5 Longer term changes

One final area worth understanding is the behavioral impacts that a maintenance project may have in the long term. Because some people may exhibit unusual travel habits while the maintenance project is going on, it is possible that once the project ends, some of those travel habits may remain. For instance, if a commuter chooses to drive his auto to work while his/her neighboring station is closed to the public, then he/she may wish to continue using it when that station reopens due to a newly formed habit. This is of particular relevance for the transit agency due to the now permanent nature of the ridership and revenue losses.

In order to examine the potential long term effects of the BLCEP, the ideal experiment would require having data *before* the project started and *after* it was completely done. However, the BLCEP has not been completed, so this condition can only be partially met by studying after some *stages* of the project have been finished. Specifically, the passengers around the stations Kimball, Francisco, Montrose and Addison will be examined before the station was closed and after the station was reopened to the public.

One of the limitations of this analysis is that other components of the project may have an impact, even after the stations reopened, hence, biasing the comparison to pre-project behavior. Specifically, three track operation started in April 2nd of 2007 and will finish in June of 2009, but the reopening date of the Montrose and Addison stations are

December 3rd 2007. In the case of Kimball and Francisco, the reopening date is before the start of Three track, which makes a case for a better comparison.

Station reopening

	Time *before* and *after*			Stations		Control zone		Difference	
Station	Before	After	Weeks	N *	%UNK**	N *	%UNK**	Dif	t-stat
Kimball	Sept-06	Feb-07	20	79	26.58%	1120	26.07%	0.51%	-0.10
Francisco	Sept-06	Mar-07	24	150	26.00%	1120	31.16%	-5.16%	1.29
Montrose and Addison	Nov-06	Dec-07	52	452	47.35%	1096	48.81%	-1.47%	0.53

* N is the number of cards that were recorded in the period before the station closure
** %UNK is the percentage of cards that did not report any activity after the station was reopened

Table 40: Long term comparison of card activity for stations and control zone

Table 40 presents a comparison of the number of cards that used the system on each of the stations *before* the closure and *after* the reopening. The data for the *after* period is collected from the same datasets of Table 21. It represents the activity of customers 2 to 4 weeks after the station reopening and is separated from the *before* data by 20, 24 and 52 weeks depending on the case.

As it can be seen in Table 40, there is a large decline in the number of cards after 20, 24 and 52 weeks. This is consistent with previous findings shown in Table 33, which showed 7-8% declines for periods of four weeks.

Comparing the percentage of declines of the BLCEP to those of the Control Zones, for the same time periods, some differences can be observed. However, none of these differences are statistically significant at a 95% confidence level.

These findings would suggest that the longer term passenger losses are negligible but the high rate of lost cards per week is very high, undermining the credibility of the method. In other words: The percentage of cards that did not report any activity consists of cards from two groups; Passengers that left the transit system and; passengers that did *not* leave the system but did *not* use their cards for other reasons. As these percentages increases over time, the portion of cards that belong to the second group will likely become larger with respect to all the group of unknown cards. Hence, the difference of the percentages of unknown cards is more likely to reflect differences in the rate of card losses than differences in the rate of transit ridership.

Both the high rate of weekly card losses and the fact that the project is not completely finished present barriers to produce an accurate picture of the project-associated long term impacts.

6 Analytical model for commuter mode change

So far, the findings presented in Chapter 5 have described the observed changes for different aspects of commuter's behavior. The results of each aspect have different degrees of certainty due to the type of data that has been analyzed. For instance, the bus modal choice is a directly observed change of behavior at the individual level and has a higher degree of certainty than other aspects like the long term behavioral changes where the conclusions are largely based on aggregate statistics.

However, the question of how the can CTA learn from these findings and be better prepared for future projects remains open. To try to answer this question we will examine commuter behavior in more detail.

This research uses the framework of random utility to explain individual choice behavior. This approach was formalized by Manski[36] (1977) and assumes that, under a set of choice alternatives, individuals will select the one that offers the highest utility. But since the utilities are not perfectly known to the analyst, they are treated as random variables. Hence, the choice that an individual selects alternative *i* is equal to the probability that the utility of alternative *i* is greater than the utilities of the other alternatives (Ben-Akiva and Lerman, 1985)[37]

$$P(i|C_n) = \Pr[U_{in} \geq U_{jn}, all\ j \in C_n]$$

Where:
Cn is the choice set of j alternatives

Equation 5: Choice probability under a random utility model

Among the behavioral components that were studied in this research, only the change of modal selection will be modeled. The chief reason for this decision is that we have access to individual responses for the mode choice decisions thanks to the AFC-AVL

[36] Manski, C. 1977 The Structure of Random Utility Models
[37] Ben Akiva M. and Lerman S. 1985 Discrete Choice Analysis

data availability which facilitates the calibration of a disaggregate model. Moreover, preliminary statistics show results according to our expectations (high rail dependence, moderate shifts to bus). However, in the case of other behavior components, our conclusions have been drawn from aggregate comparisons against a control zone.

6.1 Binary choice within a multinomial context

One special case in choice theory, when the choice set is limited to two alternatives, is called *binary.* The probability of an individual selecting alternative *i* is equal to

$$P(i) = \Pr\,(U_{in} \geq U_{jn})$$

Equation 6

Where the utility of alternative *i* is

$$U_i = V_i + \varepsilon_i$$

V_i is the systematic component and,
e_i is the random component
Equation 7

And the probability of selecting alternative *j* is

$$P(j) = 1 - P(i)$$

Equation 8

As shown in Figure 27, a rail infrastructure maintenance project is likely to affect rail passengers by offering them a choice between different alternatives: Rail, bus and other modes. The *new* rail alternative provides a lower level of service due to the inconveniences of the project while the rest of the alternatives are assumed to remain constant. Therefore, this incremental decrease can be modeled to produce forecasts of modal shifts in events of future maintenance projects.

Strictly speaking, the described situation could be modeled under a multinomial approach, where the attributes of all the possible modal alternatives (rail, bus, car, walk) are included in the model. However, the AFC-AVL data reveals the mode choice *only* from passengers that continue using the transit system. There is no direct information on the behavior of those who left the system, although the aggregate analysis made in section 5.1 shows that it was a rather small percentage. For instance, it shows that 3.9% of Three track operation's affected commuters were estimated as of having abandoned transit, but it also shows that a negligible percentage of customers affected by station closures abandoned transit. These estimations have a larger degree of uncertainty because are the result of a comparison to an imperfect control zone. Figure 28 illustrates the case of Three track operation.

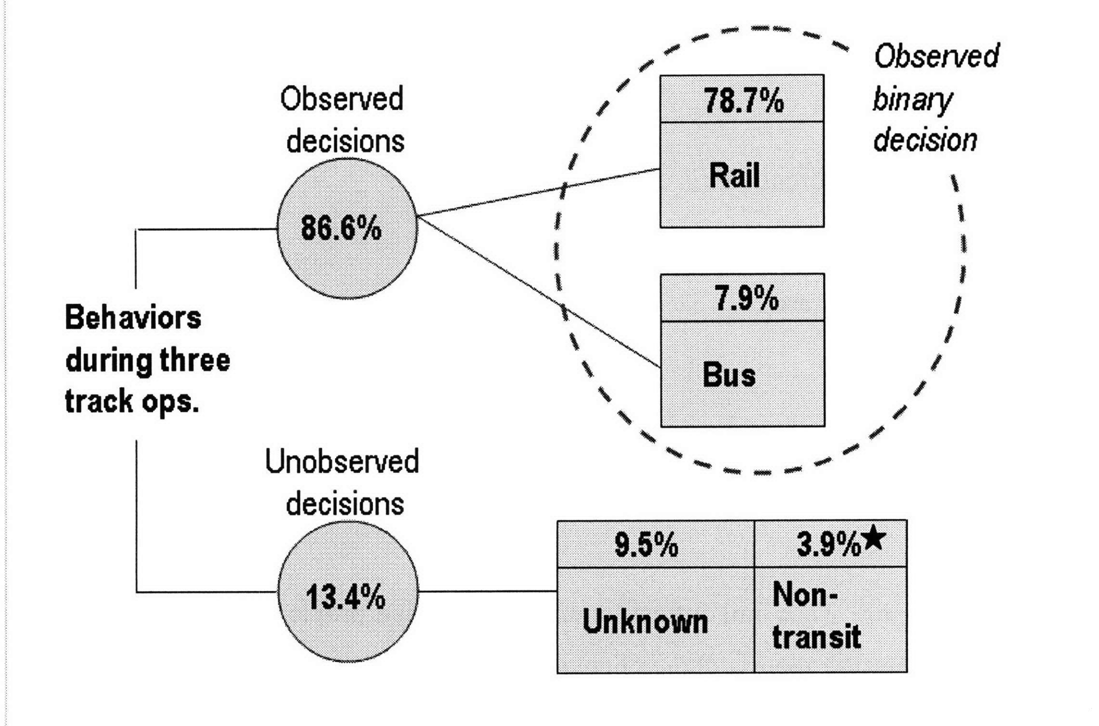


Figure 28: Scheme of observed and unobserved decision making for three track operation

★ Estimated by comparing to a Control Zone

The individual analysis for transit users' choice provides an appropriate setting to model the binary choice between bus and rail. However, the random utility modeling framework requires knowledge of the trip attributes and user characteristics that may influence mode choice. In the case of a rail maintenance project, the decision to stop using the train and start using the bus will likely be dictated by the quality of the bus service when compared to the downgraded rail service. However, the AFC data does not offer complete information about the characteristics of both alternatives. Information like how long does each user has to walk/wait/stay in the train/bus in order to reach his destination is not obvious given the complex transit system of the CTA. In order to overcome this limitation, a simulation in a network computer model was used to estimate the trip attributes.

6.2 Network Model

This research builds on previous work by Busby (2004) [38] who utilized a transit network model for the CTA to simulate impacts of system level changes on accessibility. A complete description of the characteristics of the model is documented in Busby's work. It is based on TransCAD and GIS platforms and consists of the elements shown in Table 41:

Number of Elements	Name	Driving speed	Transit speed	Walk	Drive
3048	Expressways	55	44	N	Y
4443	Major Arterials	45	36	N	Y
6913	Minor Arterials	40	32	Y	Y
2287	Ramps	40	32	Y	Y
81885	Local Streets	30	24	Y	Y
151	CTA Rail Lines	N/A	Variable	N	N
210	CTA Station Access	N/A	N/A	Y	N
265	Metra Station Access	N/A	N/A	Y	N
19698	Centroid Links	20	N/A	Y	Y
460	Metra Rail Lines	N/A	Variable	N	N

Table 41: Layers and properties that compose the network model

The model includes specific attributes about the operational characteristics of the rail and bus systems, such as: Route name and ID, Headway in the AM Peak, PM Peak, Midday, and Owl periods, Dwell time, Fare, Maximum and minimum wait time and Layover time

[38] Busby J. (2004) Accessibility based transit planning

at terminal points. The source data for these attributes is the CTA scheduled station-to-station times and other system specific information.

The original functionality of the model was aimed to gauge changes in accessibility at a regional scale. However, this research aims to understand changes in travel behavior at the individual level. This requires that each of the Smart Cards must be spatially located and integrated to the network model to calculate their transportation options and attributes. In order to accomplish this integration, the point-geographical files that compose each cards' origin and each cards' destination are added to the road network layer using the TransCad command *connect.* [39]

Figure 29 illustrates how the Smart Cards are integrated by joining reported geographic coordinates (man-like shapes) with road connectors to the road/transit network. The same procedure was also applied for the estimated destination locations in order to connect origins and destinations within the same geographic file. Those trips where a destination was not successfully inferred, were removed from the sample.

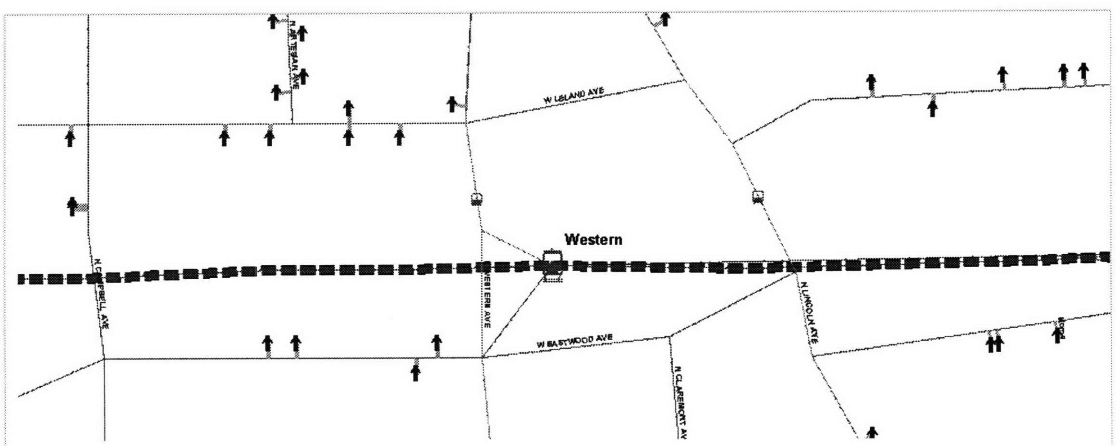


Figure 29: Addition of the Smart Cards to the road network: Snapshot around Brown Line Western station

[39] As a reference for potential future users, it is important to attach the original smart card ID to a new node in the network, otherwise there will be no record of where each smart card is.

6.3 Path choice

The network model can compute the trip attributes for an OD pair. Once all ODs are connected to the road/transit networks, their respective trip attributes were computed for both the rail and bus alternatives. The shortest path method assigns one rail path choice and one bus path choice for each OD. It selects the path with the lowest transportation cost for a set of path choices.

One of the inputs that the shortest path method requires is a set of weights for the different time components of the trip. Because people will value differently the time walking, waiting time or in the vehicle, it is necessary to compute a weighted measure of travel time. The selected weights correspond to those adopted in the Service Plan of the Toronto Transit Commission (TTC)[40]. These weights were adopted because they are utilized to evaluate the impact of changes in services on ridership, which may represent our case better than generic coefficients. Therefore, we adopted the measure of Weighted Travel Time (WTT) is the following :

$$WTT = A * invehicle + B * wait + C * walk + D * transfers$$
$$A = 1.0$$
$$B = 1.5$$
$$C = 2.0$$
$$D = 10.0$$

Equation 9: Weighted travel time: Shortest path method

Using the TransCad command *skimming*, the network model produces a matrix with the time components of each trip. However, this command must be used twice: First, to estimate the data for the rail alternative and second, to estimate the data for the bus alternative. In order to 'force' the customers to select a specific alternative, the available boarding options for the other alternative are disabled.

[40] TTC Transit Service Planning Process "Planning Transit Service," pp 7-10 (from the TTC report "Service Improvements for 2005.") (http://www.toronto.ca/ttc/pdf/service_improvements_2005.pdf)

In order to capture the impact of the maintenance project, the *before* and *during* situations were recreated in the model in the following way:

- Case 1: Rail station closure. The *skimming* command was used twice: First using all the rail stations and then disabling the affected rail stations. This way the command is forces the commuter to board the closest station and imposes the corresponding additional walking time.
- Case 2: Three track operations. The *skimming* command was used with two different train schedules: The winter 2006 (*before*) and the spring 2006 (*during*) schedules. Using this approach, additional wait time and additional in-vehicle travel time is imposed on the trips, based on the posted schedule times. The shortcoming of this approach is that the scheduled times may be deviated from the actual times, hence, biasing the analysis. However, the CTA currently does not have the technology to track the actual station-to-station train running times and the schedules constitute the best available indicator.

The matrix output of the *skimming* is initially produced for all the origins to all the destinations. Hence it is necessary to select the list of ODs that represent the patterns of the smart card customers and separate them from the rest of the matrix. In other words, it is necessary to separate the diagonal of the WTT matrix because this is the only data that we are interested in.

To summarize, the resulting output of the path choice procedure is a list of the trip attributes for each passenger's commute trip. These attributes are:

- Walk time
- Wait time
- In vehicle time
- No. of transfers
- Weighted travel time (as defined in Equation 9)

These attributes are presented for two alternatives; bus and rail and for two time periods: Before the project (*before*) and during the project (*during*).

6.4 Special cases

One of the limitations of inferring behavioral changes from the Smart Card activity is that there is no complete certainty on the motivation behind a modal shift. In the case of this research, observing behavioral changes before and during a downgrade in rail level of service is necessary, but not sufficient, to assure that the motivation to switch modes was the station closure.

The available data permits the partial detection of special cases where causes, other than the maintenance project, could have caused the modal shift:

- Different trip origin: It is possible that a commuter has changed his/her residency in-between the observed time periods. This can be inferred because there are observations of commuters who, in the *during* period, did not board a station near to his/her reported address. Hence, only if the walking distance in the *during* period is lower than 3 kms, the user is considered for the model estimation. In fact, 5.4% of all users did not meet this criteria and were removed from the sample. This filter also separates some commuters who may have given their cards to other people outside the 3km buffer.[41]
- Different trip destination: It is also possible, that the commuter changed his job location/destination in-between the two observed time periods. Hence if he/she reveals a modal shift, it may be caused by better bus access to his new job rather than because of the inconveniences of the BLCEP. In order to control for this, a proximity limit was placed so that if the two estimated destinations in the *before* and *during* periods are separated by more than 3 kms, the card is removed from the model

[41] Note: Since the nature of the maintenance project is finite and relatively short (6-12 months for a station closure) then it is unlikely that the project caused the change in location. Therefore this case is not explored in this research.

estimation. For this reason, 12.2% percent of all the records were removed with this control

- Unreliable destination estimation: In addition, for a user to remain in the sample, the estimation of his destination shall meet the rules described in section 4.3.1 for both the *before* and the *during* time periods. If at least one of these two destinations is not considered reliable, the card is removed from the sample. This filter removed 3.6% of the records.
- Bus commuters: As seen in section 4.2.7.2, there is a small portion of users that are bus commuters. Since the BLCEP is expected to affect only rail users, then this group is also removed from the data. This accounts for 2.0% of the records.

6.5 Model estimation

The selection process described in section 6.4 trimmed the sample to observe those passengers who were likely to be doing the same OD commute pattern in both time periods. It is this group of commuters whom we are interested in studying because they are the most likely to reveal behavioral changes as a consequence of the BLCEP.

	Time (min)	Average	Std Dev	Max	Min
	R_Walk	6.71	3.17	24.26	0.03
Rail *Before*	R_Wait	2.37	0.98	5.00	2.00
100% rail users	R_IVTT	21.43	8.44	62.36	2.15
	R_Trans	0.01	0.08	1.00	0.00
	WTT	38.41	11.47	105.45	10.94

	Time (min)	Average	Std Dev	Max	Min
	R_Walk	7.14	4.18	41.98	0.03
Rail *During*	R_Wait	2.84	1.21	6.07	2.00
91.63% rail users	R_IVTT	23.80	9.04	70.28	2.42
	R_Trans	0.01	0.08	1.00	0.00
	WTT	42.27	13.23	113.31	12.49

	Time (min)	Average	Std Dev	Max	Min
	B_Walk	6.66	4.89	27.64	0.09
Bus [42]	B_Wait	3.44	1.30	7.50	2.00
8.37% bus users *during*	B_IVTT	26.88	8.73	91.25	4.39
	B_Trans	0.09	0.29	2.00	0.00
	WTT	46.28	12.11	106.69	7.71

Table 42: Descriptive statistics for travel times before and during the project (N=2569)

Table 42 shows how the BLCEP affected the trip attributes for rail users. The commuters affected by rail station closures and three track operation were pooled in the same data base in order to reflect changes in walk, wait and in vehicle travel times. The total number of observations is 2,569 and, in general, 8.4% of them became bus users once the project started. The average increase in WTT is 4.1 minutes, meaning that trips were on average 9.7% longer with respect to the conditions before the project.

So, what influences a rail commuter to switch modes? According to theory, the more competitive the alternative mode is with respect to rail, the more likely that it will attract new users. In the case of the BLCEP, one way to explore the relationship between the existing bus service and the likelihood of switching modes is by comparing thetravel parameters of users who switched into a competitive bus service with those who did not.

[42] Note: No bus users *before* the project

A metric developed to compare Bus Against Rail (BAR) measures how well does the bus serve the particular OD pair for an individual passenger. The BAR is equal to:

$$BAR = WTT\_bus - WTT\_rail$$

where the WTT is calculated using the constants abovementioned in section 6.3.

Figure 30 shows the results of calculating the BAR for all the 2,569 passengers that comprise the study sample. Passengers' BAR score is sorted and plotted in an ascending order. It shows how, for the majority of the passengers (73.8%), the rail alternative turns out to be better alternative relative the bus, even after accounting for the reduced level of service due to the maintenance project.

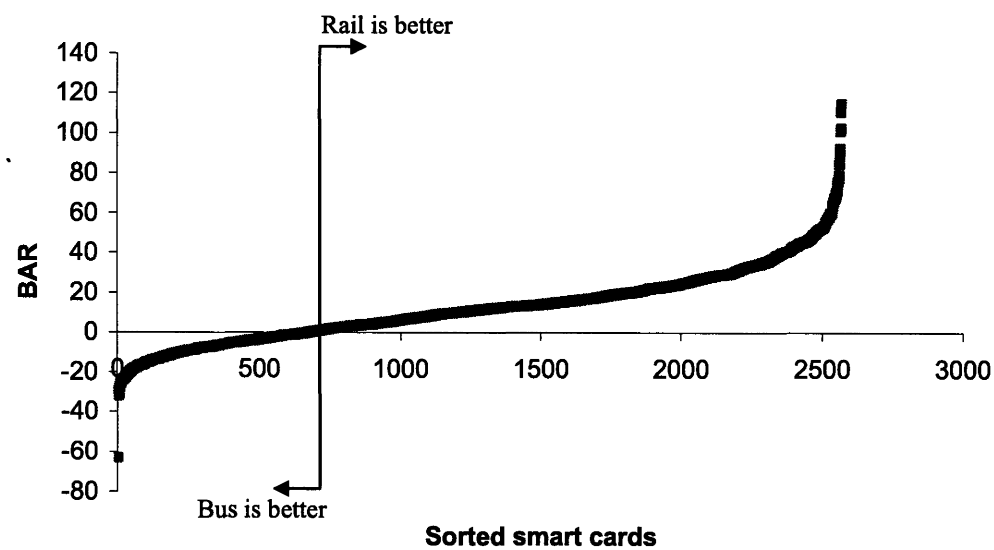


Figure 30: BAR score for passengers once BLCEP started

Passengers were divided in six groups according to their BAR score. Table 43 shows clearly how the passengers with a BAR score favorable to the bus alternative were more likely to shift modes. In the first two BAR groups, the shift rate was 21.2% and 11.8% as opposed to the remaining four groups where the average shift rate was 6.0%. When using a difference of proportions test, the first two groups (where bus is a better alternative) are statistically different than the following one (i.e. 21.2% is different than 11.8% and 11.8% is different than 5.4%). However, the shift rates for the groups where rail is better are not statistically different from each other.

BAR score	# users	%shift	Stat. Dif?
Less than -10 min	222	**21.20%**	Yes
-10 to 0 min	450	**10.96%**	Yes
0 - 10 min worse	522	5.4%	No
10 - 20 min worse	628	6.66%	No
20 - 30 min worse	364	6.06%	No
30 or more min worse	383	5.60%	--

Table 43: BAR scores compared to modal choice
*Tested at a 95% confidence level

The previous analysis showed how good bus service is linked to the propensity of passengers to use it in the event of a maintenance project. In order to quantify how the different trip attributes of each alternative influence the likelihood of switching modes, a binary logit model is developed with the collected data. The specification of the deterministic part of the utility function for each alternative is defined as:

$$Vbus = \beta wait\_bus \times Wait\_time + \beta walk \times Walk\_time + \beta trans \times No.transf + \beta ivtt\_bus \times In\_veh.time$$
$$Vrail = ASC + \beta wait\_rail \times Wait\_time + \beta walk \times Walk\_time + \beta trans \times No.transf + \beta ivtt\_rail \times In\_veh.time$$

Where each β coefficient represents the marginal disutility of investing one unit of time/transfers. Waiting and in-vehicle were assigned a separate coefficient for each alternative in order to represent different travel conditions across transit. ASC is the alternative specific coefficient that represents the inherent preference for rail, all things equal.

Table 44 shows the results of the model estimation[43]: After using maximum likelihood estimation techniques, it is seen that the coefficients have an expected negative sign for all the time attributes, supporting the notion that commuters will dislike an alternative as the invested time increases. All coefficients are all significant with a 90% confidence level, except for β Rail_wait time, which is still significant at an 84% confidence level. This can be explained by the lack of variability across rail wait times, as only two lines were examined with similar AM peak headways

Name	Value	Robust Std err	Robust t-test	p-value
ASC_RAIL	1.64	0.491	3.33	0.00
BBUS_IVTT	-0.0528	0.0127	-4.15	0.00
BBUS_WAIT	-0.133	0.0706	-1.88	0.06
BETA_TRANS	-0.814	0.408	-1.99	0.05
BETA_WALK	-0.0970	0.0162	-5.98	0.00
BRAIL_IVTT	-0.0306	0.0104	-2.94	0.00
BRAIL_WAIT	-0.0883	0.0632	-1.40	0.16

Number of observations:	2569
Final log-likelihood:	-702.606
Likelihood ratio test:	2156.178
Rho-square:	0.605
Adjusted rho-square:	0.602

Table 44: Model estimation results

The magnitudes of the coefficients reveal how passengers are willing to trade off different attributes. In order to estimate the extent of these trade-offs, the Marginal Rate of Substitution (MRS) is calculated for each attribute with respect to the in-vehicle travel time.

In the case of the rail wait time, the MRS is equal to:

43 Other model specifications were tested using different coefficients for bus and rail but no significant improvement was found, either in model fit or coefficient estimation.

$$MRS = \frac{\dfrac{\partial U}{\partial wait\_rail}}{\dfrac{\partial U}{\partial inveh\_rail}} = \frac{-0.085}{-0.033} = 2.89$$

It can be interpreted that rail commuters are willing to trade off up to 2.6 minutes of in-vehicle time for one minute of wait time. This is expected because the waiting time can be perceived as 'avoidable' if the user knew the exact arrival of the train, as opposed to the time in the vehicle which is a fixed amount of time for every day and is perceived as a 'budgeted' time.

In the case of the walk time, the MRS is equal to:

$$MRS = \frac{\dfrac{\partial U}{\partial walk}}{\dfrac{\partial U}{\partial inveh\_rail}} = \frac{-0.076}{-0.033} = 3.23$$

The walk time is also perceived more negatively than the rail in-vehicle time, at a ratio of 3.2. This is also expected because the walk time can be made under adverse weather/walking conditions and is imposing the commuter to use his energy to reach a destination, as opposed to the in-vehicle time where the role of the passenger is rather passive.

The alternative specific coefficient for the rail alternative is also of interest. Its positive sign shows the natural preference in favor of rail based alternatives. Moreover, its high magnitude, relative to the time coefficients, is expected because the sample is composed by rail commuters who had a previous habit of using the train.

When comparing across modes, it can also be seen the preference of passengers to wait in a rail station as opposed to a bus stop. The MRS is equal to 1.59 and reflects the more favorable conditions of waiting in a rail station:

$$MRS = \frac{\dfrac{\partial U}{\partial bus\_wait}}{\dfrac{\partial U}{\partial rail\_wait}} = \frac{-0.133}{-0.0883} = 1.59$$

Finally, the transfer coefficient is very large with respect to the in-vehicle travel time coefficient. This is showing how a linked trip (one that includes a transfer) will not likely be a good substitute for a direct trip and will not be attractive to foster a behavioral change.

The MRS for the different trip attributes with respect to the in-vehicle travel time could be regarded as new weights in the event of estimating a new Weighted Travel Time (WTT) for each alternative. The differences in magnitudes, compared to those of the TTC, could be caused by the fact that this research is only examining at rail commuters, while the TTC's approach is for a broader base of users. These new weights, however, have the value of being estimated for Chicago-based rail customers and better represent the local conditions of waiting and waiting.

Other specifications were tested as well, but there was no significant improvement in the model's fit. Attempts to include an income variable in the model, based on the household income of each neighborhood, also decreased the fit of the model. Future modeling attempts can be greatly improved by associating a Smart Card holder with his/her particular socio economic characteristics and his/her other travel alternatives. This would require an effort to survey the current customers to collect additional information and can be explored as an area for future research.

6.6 Application to passenger forecasting

One application of the resulting model just described in section 6 is the ability to forecast modal shifts for passengers that will be subject to the degradation of the rail level of service. This model applies to cases such as station closures, reduction in train frequencies, decrease in train speeds, etc. Passenger forecasting is an essential step in the process of planning supplementary bus routes and estimating potential revenue losses.

The main application of this binary model is in the form of an incremental logit. According to Ben-Akiva and Lerman (1985)[44] this form can be used to *"predict changes in behavior on the basis of the existing choice probabilities of the alternatives and the changes in variables that obviates the need to use the full set of independent variables to calculate the choice probabilities"*. In other words, forecasting with the current model requires to have knowledge of the magnitude of the utilities before the project and the extent of the downgrade in level of service, for each affected individual.

The choice probability derived from a change in utilities (penalties imposed by the project) is calculated with the following equation:

$$P_n(i)^* = \frac{P_n(i)e^{\Delta V_{in}}}{\sum_{j \in C_n} P_n(j)e^{\Delta V_{jn}}}$$

Where:
$P_n(i)$ is the choice probability before the reduction in level of service[45]
$P_n(i)^*$ is the new choice probability that the individual n will choose alternative i

ΔV_{in} is the change in utility for individual n and alternative i, $\sum_{k}^{K} \beta_k \Delta x_{ink}$

Equation 10: Choice probability for an incremental logit

Therefore, the share of passengers that shift from rail to bus will be given by the equation:

[44] Ben Akiva M. and Lerman S. 1985 Discrete Choice Analysis
[45] Note: Although in reality $P_n(i)$ is equal to 1 (because all of the customers are rail users in the first time period), the model will predict result different than 1, due to its probabilistic nature.

$$\%Shift = (1 - \%other) \times \left(P_n(rail)*\right)$$

Where:
% other is an estimate of the percentage of passengers that will not use transit (for three track operation it was an average of 3.9%)
Equation 11: % of passengers that will shift from rail to bus

And, in turn, the share of passengers that will continue using rail is defined as:

$$\%Stay = (1 - \%other) \times \left(1 - \%Shift\right)$$

Where:
% other is an estimate of the percentage of passengers that will not use transit (for three track operation it was an average of 3.9%)
Equation 12: % of passengers that will stay using rail

In order to illustrate the use of the model in forecasting, we explored a hypothetical situation of transportation planning in Chicago is here described: The Fullerton station is closed to the public and announced with time in advance. What level of supplementary bus service should be provided in the AM peak?

I order to answer that question, we first answer the following one: What percentage of users would we expect to leave the transit system? According to the results in Table 33, there is no statistical evidence that shows a reduction in transit users after any of the four station closures. However, an estimated 3.9% of the users affected by three track operation used non-transit modes for their trips. In order to overcome this uncertainty, a conservative estimate of 2.5% will be used to represent the percentage of passengers that will not use the CTA.

The next step consists of answering the following questions: What percentage of the users will switch to bus? What percentage will use rail by walking to the next station? Two approaches are proposed to answer these questions.

6.6.1 Disaggregate forecast

This approach evaluates the changes in choice probabilities for each individual based on the impact that the project has on each particular commuter. This approach likely yields better results but is very time consuming and requires access to a GIS-based transit network:

- First, the process described in section 4.2 and represented in Figure 11 should be followed by geo-coding the rail commuters that will likely be affected by the station closure.
- Second, the process described in section 4.3.1 should be followed to infer the destinations of the affected commuters.
- Third, the trip attributes for each of the two transit alternatives should be estimated by using the CTA network model. The process in section 6.3 describes the commands that are appropriate for this procedure.
- Fourth, the choice probabilities for both alternatives must be computed by using the following equation:

$$P(i) = \frac{e^{V_{in}}}{\sum_{j \in C_n} ^{V_{jn}}}$$

- Fifth, the effect of the station closure on walk times must be computed for each individual. This is done by using the procedure described in section 6.3 Case 1 by estimating the additional walking time for each passenger. The respective induced disutility is calculated with the formula:

$$\Delta V_n(rail) = \beta_{walk} \times \Delta walk_n$$

- Sixth, the new choice probabilities are calculated by applying Equation 10 for each individual. As explained, this equation only requires the original choice probability and the change in utility for each individual.
- Seventh, the new shares of rail and bus are calculated using Equation 11 and Equation 12

Intermediate results of this procedure are shown in Table 45:

	Time (min)	Average	Std Dev	Max	Min
	R_Walk	7.567	2.536	13.960	0.920
Rail	R_Wait	2.443	1.066	5.000	2.000
	R_IVTT	14.580	4.319	43.360	6.180
	R_Trans	0.000	0.000	0.000	0.000
	B_Walk	6.089	3.732	17.510	0.090
Bus	B_Wait	4.218	1.167	6.000	2.500
	B_IVTT	20.839	3.786	54.600	9.460
	B_Trans	0.004	0.060	1.000	0.000
Imposed penalty of station closure	*R_Walk*	*3.749*	*3.045*	*9.904*	*0.000*

Table 45: Descriptive statistics for travel times in Fullerton (N=275)

The new modal shares are presented in Table 46.

	Rail	Bus	Other
Before station closure	100%	-	-
During station closure	87.2%	10.3%	2.5%

Table 46: Resulting modal shares of Fullerton closure

6.6.2 Aggregate forecast

The second proposed approach to forecasting is based on aggregate information at the station level. This approach is based on the concept of aggregate logit elasticity which represents the *"responsiveness of some group of decision makers rather than that of any individual"* Ben-Akiva (1985). In other words, mode shift forecast is made for all the group of commuters of a specific station at once, instead of evaluating individual changes in choice probabilities.

Using this approach requires the computation of $\overline{P}(rail)$ as the expected share of the group of individuals before any change in the level of service. It is equal to:

$$\overline{P}(rail) = \sum_{n}^{N} \frac{P_n(rail)}{N}$$

Equation 13

As pointed out before, $\overline{P}(rail)$ is 1 in reality because all the sample is composed by rail users, but the model will output a figure that is less than 1 due to the probabilistic nature of the logit model. Now, in order to calculate the elasticity of the *group* to a change in attribute *k*, the following formula must be used:

$$E_{Xjnk}^{\overline{P}(rail)} = \frac{\beta_k}{N \cdot \overline{P}(rail)} \cdot \sum_{n=1}^{N} P_n(rail)[1 - P_n(rail)]x_{jnk}$$

Where the elasticity $E_{Xjnk}^{\overline{P}(rail)}$ represents the incremental change in $\overline{P}$ as a response to an incremental change in an attribute k

Equation 14

The aggregate forecasting method is better employed when the increment is constant across commuters. For instance, if we want to forecast the number of passengers that will shift to bus given that trains will be 10% slower. On the other hand, if the impact varies substantially across individuals (like in a station closure), then the disaggregate method will attain better accuracy.

With the observed data, we generated elasticity figures for walk time and in vehicle travel time as shown in Table 47. These elasticities can be interpreted as the percentual decrease in the probability to use rail as a consequence of increasing that attribute by 1%. Indeed, these are cross-elasticities as well because they are estimated within a binary model, meaning that the decrease in rail share should be interpreted as an increase in bus share. These elasticities can vary by area of the city, depending on the quality of the local bus service, but represent average values for the North Side of Chicago. Future planning efforts can calculate elasticities of specific stations to create out-of-the-pocket figures to estimate ridership impacts given changes in specific trip attributes

Attribute	Elasticity
Walk time	-0.0587
Wait time	-0.0192
In vehicle time	-0.0561

Table 47: Aggregate elasticities for rail to bus shift

These walk time elasticities could be used to recalculate the results of the hypothetical situation of the Fullerton station closure. Differences in bus shares between the disaggregate and an aggregate forecasts can reflect the incurred inaccuracies of assuming a constant penalty value. In this case, an average additional walking time is assigned to all users, while in reality, each user will have its own unique impact. The disaggregate approach is recommended for more accurate forecasts but it requires a more detailed understanding of each passenger's trip patterns and alternatives, hence, the usefulness of the elasticity figures.

6.7 Revenue implications and social costs

The revenue implications of a station closure can be derived from the forecast. Passengers who leave the transit system means lost revenue while passengers who switch to bus will pay a lower fare. Using an estimate of 7,500 daily boardings in the AM peak, Table 48 shows that the monthly revenue losses are modest ($12,499) and are driven mostly by passengers who leave the system. In the event of a 12 month station closure, the total revenue losses become more significant ($149.985). These figures are useful for the standpoint of financial analysis and can be used to gauge different engineering strategies for maintenance projects

AM peak impact	Passengers			Revenue losses		
	Rail	Bus	Non transit	Shift to bus	Leaves CTA	Total loss
Per working day	6540	772.5	187.5	(193)	(375)	(568)
Per month	143880	16995	4125	(4,249)	(8,250)	(12,499)
For a 12 month project	1726560	203940	49500	(50,985)	(99,000)	(149,985)

Table 48: Revenue losses derived from mode shifts in the event of a Fullerton Station closure

Another cost derived from the mode shift is the provision of supplementary bus service. This cost calculation will depend largely on the existing capacity of the bus routes and the characteristics of the existing routes. In the spirit of estimating an order-of-magnitude value of this cost, a simple calculation is here presented. It assumes that the 10% of the "new" bus users can be served with the existing capacity. However, the remaining 90% must be served by scheduling additional runs. Average CTA cost figures are used in this calculation for revenue-bus-hour ($34.83) and revenue-bus-miles ($3.06)[46]. Other bus and route characteristics are presented in Table 49. The sum of the revenue losses and the cost of supplementary bus service adds up to $1,094,445 in a 12 month span. This cost figure is not insignificant in the context of the Brown Line Capacity Expansion Project, where the total cost *per station* is close to $29.4 Million dollars[47].

[46] CTA operational cost model 2007, not published
[47] The total cost of BLCEP was 530 million dollars involving a total of 18 stations

AM peak impact	New bus passengers *	Necessary supplementary runs	Cost revenue bus hours ($)	Cost revenue bus miles ($)	Total operational cost ($)
Per working day	695.25	20	(3,483)	(3,672)	(7,155)
Per month	15295.5	440	(76,626)	(80,784)	(157,410)
For a 12 month station closure	91773	2640	(459,756)	(484,704)	(944,460)

Other assumptions:
Max passengers per bus: 60
Bus cycle time (min): 100
Duration AM peak (hrs) 3
Roundtrip Route Length (mi) 20
*10% of the passengers that shift to bus are assumed to be captured by existing capacity.

Table 49: Approximate cost of providing bus supplementary service in the event of the Fullerton station closure

Figures in Table 49 will change by assuming a higher number of the new passengers being absorbed by existing capacity. As mentioned, this example used 10% , but it is likely to across different stations depending on the crowding of nearby bus routes. Table 50 shows that the total cost of bus provision can range between $613.000 and $944.000 for a 12 month project

% absorbed by existing capacity	Cost of bus provision for a 12 month project ($)
10%	(944,460)
25%	(802,791)
40%	(613,899)

Table 50: Sensitivity analysis for cost of additional bus provision

A partial social cost analysis can also be performed for the case of the Fullerton station closure. Using station averages from Table 45, it can be seen that the average bus trip is approximately 6.5 minutes longer than the average rail trip. Also, it can be seen that the average imposed walking time for those who decide to continue using rail is 3.75 minutes. This means that for those passengers who decide to continue using transit there will be an approximate daily social cost equal to:

$$6{,}540 \times 3.75 + 772 \times 6.5 = 29{,}543 \text{ min}$$

This cost in minutes can be converted into a monetary value for evaluation purposes. This requires to use a constant unit for the Value of Time, which can vary across, trip purpose and magnitude of the impact (AASHTO ,1977)[48]. The impacts associated with the station closure have an average 3.75 minutes of additional walk time and can be as high as 10 minutes.

Therefore, we used figures between \$3.50 and \$11.50 to represent the value of time[49]. The equivalent monetized daily cost varies between US\$1,700 and US\$5,600 in the AM peak. Assuming that each commuter will have an equal imposed cost on the P.M. trip, the monthly social cost will range between US\$55,000 and \$226,000. Finally, a 12 month project will impose a social cost ranging between US\$660,000 and US\$2,715,000. This is an under-estimate because it does not account for the imposed costs on those who will not use transit. However it presents a good idea of the order of magnitude of the social costs imposed by a sample station closure.

These analyses speak of the relevance of providing a good level of supplementary service and planning adequately the engineering stages of the project: Underestimating the level of supplementary buses will likely turn into overcrowded buses and unsatisfied customers. On the other hand, overestimating the amount of supplementary service will impact the finances of the project in an important fashion. Large engineering projects should budget the inconveniences that are caused to the customers and plan for optimal strategies that balance the direct costs of civil works and the imposed costs on the customers and the agency with the future benefits of each different alternative.

[48] AASHTO (1977). Manual on User Benefit Analysis of highways and Bus-Transit Improvements.
[49] Estimated based on a \$48.201 median household income. US Census bureau. Current Population Survey, 2006 http://pubdb3.census.gov/macro/032007/hhinc/new04_001.htm

7 Conclusions

7.1 General conclusions

a. Smart cards have become an important fare media for many transit agencies in the world. The CTA has introduced these devices in the form of the Chicago Card (CC) and the Chicago Card Plus (CC+). However, the rate of usage of the Smart Cards is rather low, compared to other agencies in the world, despite recent incentives to increase its usage. Although it is not clear why the Smart Cards are not more popular, it is likely related to the inability to provide monthly and weekly pass options in the CC form, the strong inclination of the device to be a better fit for rail use and under-perceived benefits of using it on some segments of the population. Smart Cards, as a data source, open an important room for transit planning applications which is currently under exploration by researchers and the agency itself. This thesis studied the case of the BLCEP as a research application in travel behavioral changes by studying a maintenance project in Chicago with the Smart Cards as data source

b. Transit infrastructure in Chicago requires significant inflows of capital in order to bring it to a "state of good repair". Maintenance efforts are at risk of being neglected due to funding constraints and recently related incidents have shed light on the need to properly maintain the infrastructure. The Brown Line Capacity Expansion Project (BLCEP) is an example of a staged approach to infrastructure maintenance which aimed to combine significant construction works while reducing impacts on passengers. After analyzing boarding counts at the stations it was found that the BLCEP induced a reduction in rail ridership at the line and at the station level. This analysis was a good starting point to quantify changes in travel behavior but opened the door to a more fine grained analysis at the individual level.

c. Using the Smart Cards registration data in conjunction with AFC and AVL data is a useful way to analyze individual trip patterns. By following a step-by-step methodology it is possible to create user profiles that describe passengers habits while overcoming some of the natural limitations of using entry tap-only data. The methodology presented was applied for weekday commute trips but can be generalized for other days and times. One caveat of relying on Smart Card data is the inability to know more characteristics of the individual and the limitations in sample sizes and ticket choices. The methodology was successfully applied to two time cross-sections around the BLCEP and proved successful by turning into a pseudo-panel database of customer activity.

d. The analysis of individual customer activity allows an analyst to quantify and evaluate changes in travel behavior by comparing two time periods. In the case of the BLCEP, changes were detected in bus modal selection and boarding locations and, to a lesser extent, in non-transit modal selection, trip frequency and boarding time. A key barrier to make accurate analyses is the high rate of random card losses. One major advantage of these results is that they can be explained by developing econometric models at a very low cost with the Smart Card data compared to a full passenger survey.

e. The random utility framework is useful to model changes in mode selection and proved appropriate to model rail to bus modal shifts. Exploration of the data shows that the decision to use the bus under the event of a maintenance project largely depends on the quality of the competing bus service. The model estimation also shows how customers have lower values of their trip time while in the vehicle as compared to walking or waiting. Applications of this model show that the costs associated with providing adequate bus level of service are not negligible from the societal and the financial points of view, hence, the importance of using quantitative techniques to plan for maintenance events.

7.2 Specific conclusions

a. This thesis developed a replicable methodology to study different aspects of travel behavior by examining Smart Card activity. It is the continuation of a trend of research pieces for the CTA and presents two main methodological contributions:
(i) the analysis of two time periods and,
(ii) the use of a network model to estimate trip attributes.
This type of analysis can be replicated to test the effect of different operational and fare policies and marketing strategies over time. However, two main barriers still remain: The first one is the relatively low penetration of the Smart Cards within the city, as some areas do not have a good coverage and the bus service still lags in market share with respect to rail. The second one is the high rate of card losses that was observed, which can significantly trim a panel of observations over time. Further analyses and research designs should take these aspects into account

b. The examination of commuter patterns with the Smart Cards presented practical information useful for planning and evaluation purposes.
(i) The seasonal fluctuations of Smart Card holders is not as pronounced compared to the boardings in the rest of the system. This reinforces the hypothesis that Smart Card holders represent a base of transit frequent users.
(ii) Smart Card boardings show that, in average, a weekday carries 17.6% of the weekly boardings, being Wednesday the busiest day of all. The weekend carries 12.7% of all boardings. As practical ratios, it can be said that Saturdays carry 43.5% and Sundays carry 28.4% of a weekday's boardings.
(iii) A time of the day analysis shows that the busies segments of the day are 7-9 A.M, where 26% of all boardings are recorded and 5-6 P.M., where 14.2% of all boardings are recorded as well.
(iv) A walking distance analysis shows that, in average, rail commuters walk 620 meters to the analyzed stations. Also, there is a high percentage (13%) of the Cards whose reported addresses does not correspond to the current residence of the user.

(v) A weekly time of the day analysis shows that, across the boardings in the AM peak, more than 65% of them are made by passengers that use the system at least 3 times per week and 50% of them are by passengers that use the system 4 or more times per week.
(vi) Across frequent users (more than 3 trips/weeks) 85% of them showed commute patterns by boarding the system within a 30 minute window. This shows that regular commuters represent at least half of the boardings in the AM peak underscoring the relevance of a commuter-based analysis.
(vii) Examining the return trip of rail commuters also shows that, while 79% of commuters use rail for all their return trips, many of them alternate their journeys with bus boardings. The most used routes for an afternoon return trip were 151, 147 and 22. A more extended analysis to all the north side of Chicago can help to generalize this finding and determine the most used bus routes for return trips.

c. The Smart Card analysis also permitted to examine modal shifts across a group of customers. In the case of the BLCEP, two different cases were analyzed: Station closures and Three track operations. Station closures showed that 11.5% of the customers switched to the bus after four weeks of closure, while no significant amount of customers were detected leaving the system to other modes. In the case of three track operations, 7.9% of the examined customers shifted to bus and an estimated 3.9% used other modes. It can be seen that station closures were not as badly perceived as three track by customers and did not cause passenger losses to the CTA. It can be explained because the announced duration of the station closures (6-12 months) is shorter than the duration of three track operations (30 months). However, it is yet to be proved if these losses are permanent and whether the system will recover those riders once the operations go back to normality.

7.3 Recommendations

a. Provide effective supplementary services to transit commuters: Projects like the BLCEP will create inconveniences to customers and, although these will be compensated by the future benefits of the project, it is important to mitigate the temporary negative effects of these in order to prevent permanent passenger losses. The CTA's North Side commuter market showed to be a fairly captive market to transit. It responded to BLCEP by continuing, in majority, to use the train or the bus for work trips. A smaller segment presumably left the CTA for other modes, but was only noticeable during three track operations. Therefore, its important to acknowledge these captivity conditions for the planning of future projects.

b. Use analytical methods to stage a maintenance project: Current computational tools allow analysts to evaluate the outcomes of different construction alternatives in terms of ridership losses, modal shifts, and the corresponding revenue implications and social costs. This research presents a methodology and a model to improve the planning processes for maintenance projects; in particular, to aid the design of supplementary bus services. It also presents out-of-the-pocket elasticity figures that can be used to gauge the ridership implications of alternative service changes. In any case, the models established in this thesis are only applicable in the context of rail maintenance projects and minor decreases of Level of Service

c. Continue monitoring ridership fluctuations: BLCEP is a long project that is interesting to analyze from the point of view of passenger demand. This research was able to present some of the short-term ridership implications, but was unable to show long term effects. The project is, as of June/08, still undergoing and some of the monitoring techniques presented in this research can be replicated to make an updated evaluation and assess permanent passenger losses.

d. Design a Smart Card based data collection program: The use of Smart Cards for planning purposes could be improved by convincing the users –at least a group- to share additional information with the CTA about their travel patterns and personal characteristics. In this way, the agency could have a better understanding of the impact of different operational policies and strategies. This is a fully unexplored area for both research and practical purposes and could start as a pilot by querying how some of the customers abandoned the CTA when three track operation started.

7.4 Future research

This research continues a series of projects centered around the potential applications of automated data to transit planning purposes. This project, in particular, challenges common practices of data collection and aggregation: First by proposing ways in which the Smart Card records can be used as a *de facto* panel of revealed preferences to study passenger behavior. And second, by proposing and illustrating how tasks of modeling and forecasting can be made at the individual level without the need to aggregate them.

However, the results of this research have a number of caveats that can be addressed by future researchers:

First, there are concerns regarding how representative the resulting sample of smart cards is. Although the number of cards around some stations resulted to be sufficient for our purposes, the concern looms over the particular behavior of smart card holders as opposed to non-smart card holders. This concern can play a bigger role given that the passes are not popular across Smart Card holders and that pass holders have a different perception of the cost of the trip. A relevant topic for a future researcher can be to estimate the behavioral differences between Smart Card holders and non smart card holders.

Second, the results of this research can be improved by contacting those Smart Card holders that did not report activity during BLCEP. This would allow the CTA to have a better understanding of the changes in travel behavior (if any) of these customers, as these are unobserved using a Smart Card approach. This is particularly relevant because within this group there are customers who decided to use other modes for commuting and represent a passenger loss for the agency,

Third, this study was limited to commute trips. In traditional transportation planning analysis trips are classified by purpose as passengers may perceive trip attributes in a different way. Future research can be dedicated to detect other-than-work trip patterns in order to enrich the current analysis.

Fourth, the scope of a Smart Card activity analysis can be greater by looking at bus users too. This research examined rail frequent users to take full advantage of observing the impacts of the BCLEP. However, major bus route changes should also provide insights to examine passenger behavior.

These analyses could have a great deal of help by developing a software tool that helps to reduce processing times. The step-by-step procedure to select data, coupled with the TransCad skims, takes a long time to process in order to produce user profiles. This thesis was developed with months of work but actual project evaluation will not afford that much time for data collection and analysis.

8 References

Ben Akiva M. and Lerman S. 1985 Discrete Choice Analysis

Busby J. Accessibility based transit planning, Massachusetts institute of Technology, 2004

AASHTO. Manual on User Benefit Analysis of highways and Bus-Transit Improvements, 1977

Chicago Transit Authority, Fare Media Report, June 2007

Chicago Transit Authority, Market research. CTA Smart Card market: What do we know?. 2007

Chicago Transit Authority, Rail ridership report by line and branch. June 2007

Chicago Transit Authority, Rail ridership report by line and branch. December 2006

Chicago Transit Authority, Project Presentation 03/01/07. Olson Auditorium. www.ctabrownline.com/pdfs/communitymeeting_3-1-07.pdf, accessed on April 15th, 2007

Chicago Transportation Authority. Operational cost model 2007, unpublished

Chicago Transit Authority website press release 4/13/04. http://www.transitchicago.com/news/archpress.wu?action=displayarticledetail&articleid=119462

Chu, K.K./ Chapleau,R. Enriching Archived Smart Card Transaction Data for Transit Demand Modeling, Paper submitted to the 87th Annual Meeting of the Transportation Research Board, 2007

Cui, Alex Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems. MIT Master's Thesis, 2006

Gupta, Saumya. Understanding Transit Travel Behavior: Value added by Smart Cards. MST thesis, Massachusetts institute of Technology, 2006

Manski, C. 1977 The Structure of Random Utility Models

Regional Transportation Authority, Moving beyond congestion (2007)

State of Illinois website, http://www.illinois.gov/PressReleases/ShowPressRelease.cfm?RecNum=1076&SubjectID=25

TCRP 94, Fare Policies, Structures and Technologies: Update, 2003

Toronto Transportation Comission. Transit Service Planning Process "Planning Transit Service," pp 7-10

Transystems Corporation . Memorandum: CTA SmartCard Traveler Database Feasibility, 2003

Trepanier M., Tranchant N. and Chapleau R. Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System', Journal of Intelligent Transportation Systems, 11:1, 1 – 14, 2007

Utsunomiya, Attanucci and Wilson Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning, 2006

Wilson, Zhao and Rahbee. The potential impact of automated data collection systems on urban public transport planning, 2005

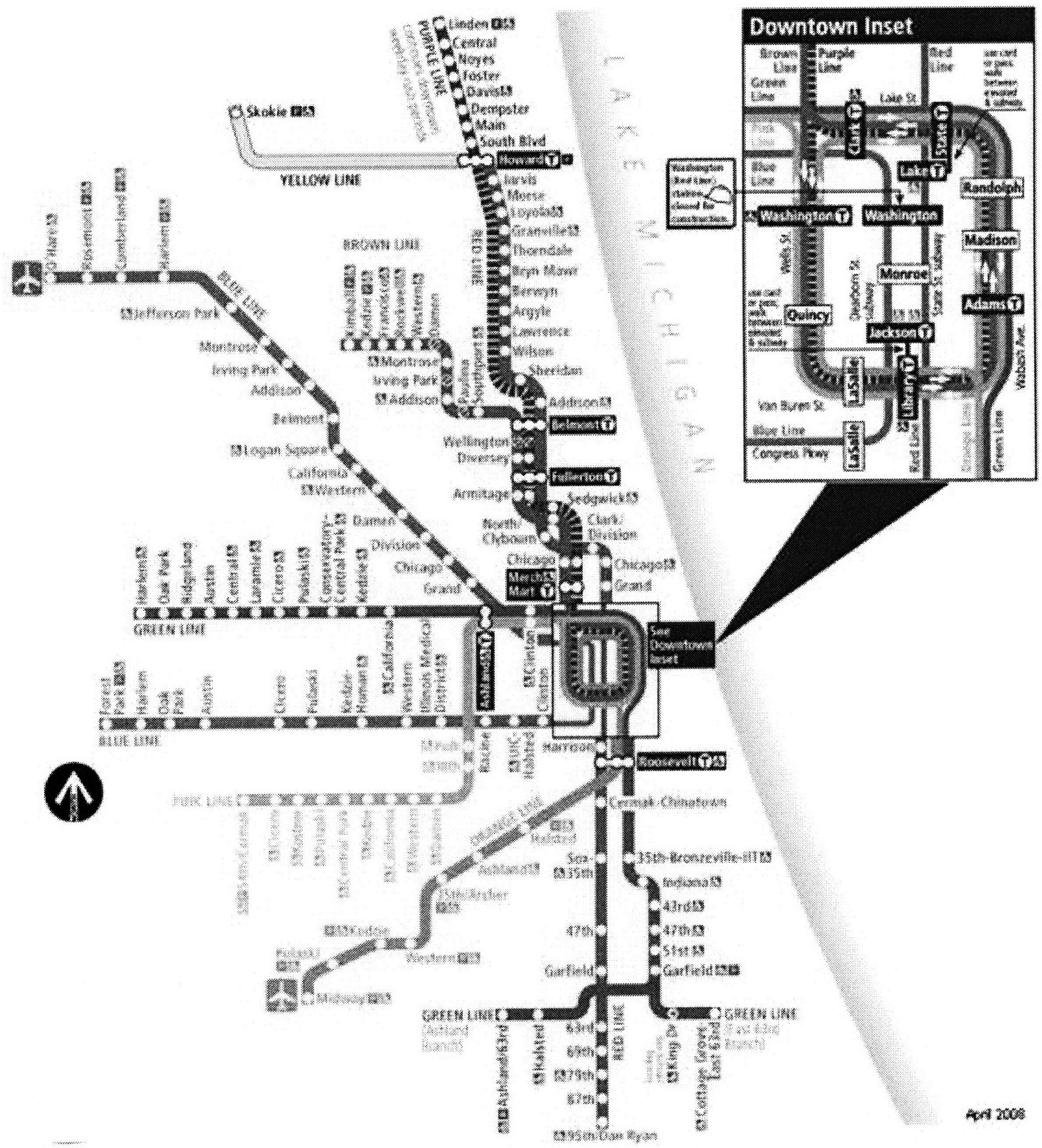
Yi, Hong Transition to Smart Card technology: How transit operators encourage the take-up of Smart Card technology. MST Thesis, Massachusetts institute of Technology, 2006

Vargas Astaiza, J. and Minser J. Expanding CTA's Smart Card Program: Market Research Efforts, 2006

Zhao, Jinhua, The Planning and Analysis Implications of Automated Data Collection Systems: Rail Transit OD Matrix Inference and Path Choice Modeling Examples, MST, MCP Thesis, Massachusetts institute of Technology, 2004.

APPENDIX

CTA Rail network map [50]



[50] Source: www.transitchicago.com. Viewed on April 28th 2008