The Role of New Technology in Improving Data Collection for Public Transportation

by

Albert Shing Fai Wong

Submitted to the Department of Civil and Environmental Engineering and the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the Degrees of

Master of Science in Transportation, Master of Engineering in Electrical Engineering and Computer Science, and Bachelor of Science in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1997

Author ....
Department of Civil and Environmental Engineering and Department of Electrical Engineering and Computer Science
February 1997

Certified by ....
Professor Nigel Wilson
Department of Civil and Environmental Engineering
Thesis Supervisor

Accepted by ....
Professor Arthur C. Smith
Chairman, Committee on Graduate Studies

Accepted by ....
Professor Joseph Sussman
Chairman, Departmental Committee on Graduate Studies

The Role of New Technology in Improving Data Collection for Public Transportation

by

Albert Shing Fai Wong

Submitted to the Department of Civil and Environmental Engineering and the Department of Electrical Engineering and Computer Science on January 27, 1997, in partial fulfillment of the requirements for the Degrees of Master of Science in Transportation, Master of Engineering in Electrical Engineering and Computer Science, and Bachelor of Science in Computer Science and Engineering

Abstract

The feasibility of relying primarily or exclusively on electronic fareboxes to collect the information needed for monitoring the performance of public transport services at the individual route level is investigated through surveying various transit agencies in the United States and testing a number of proposed methods for estimating revenue and ridership information at different levels of detail from electronic farebox data.

The surveys show that lack of driver compliance is the major obstacle to the successful use of electronic fareboxes to estimate revenue and ridership information for planning and service evaluation purposes. Agencies which are successful with electronic fareboxes are usually those with considerable experience with the electronic farebox or those with good and comprehensive training, retraining, and disciplinary programs for operators. Most transit agencies use electronic farebox revenue data for planning and service evaluation purposes, but only a few, usually the ones which are more successful with the electronic farebox, are using ridership data from the electronic farebox.

The proposed methods focus on four types of information: daily route fare revenue, daily route pass swipes, daily route ridership, and ridership by route and time period. The test data were provided by the Massachusetts Bay Transportation Authority, and both 1994 and 1996 data were used to see the effects of different driver compliance levels. The testing of the proposed methods confirms that driver compliance is very important to the success of using electronic farebox for planning and service evaluation purposes. It also shows that the electronic farebox generally works better with large routes than small routes, and that trip-level screening is essential. Furthermore, it shows that estimated daily route fare revenue is the most reliable information among the ones estimated in this study, followed by daily route pass swipes, daily route ridership, and ridership by route and time period.

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

Acknowledgements

I would like to thank all the people who lent me their time, machine, and knowledge to support my work. They really made my research smooth and enjoyable.

First, I would like to thank the University Transportation Center for funding this research and the Massachusetts Bay Transportation Authority for providing all the necessary data and information. Without the UTC or the MBTA, this research project would have been impossible.

In addition, I would like to thank Clinton Bench from the Massachusetts Bay Transportation Authority, who provided me with all the data that I needed to complete my work and surveyed some of the transit agencies in the United States for me, which formed the basis of Chapter 2. In addition, I would like to thank Alan Castaline and Lasana Kamou for their opinion, guidance, and support.

My thanks also go to Chuck Sawyer, Desi Lewis, and Martin Olsen of the Seattle Metro Transit Authority, Matt Davis, Henning Eichler, and Sandra Patton of the Greater Cleveland Regional Transit Authority, Alan Erenreich, David Jordan, Larry Hirsch, Milton Laska, Brian Cooper, Yorkman Yu, and John Kennes of the New York City Transit Authority, Leslie Stepney and Emmett Crockett of the Washington Metropolitan Area Transit Authority, Sarah LaBelle, Jim Lachowicz, Chung Chung Tam, Ross Patronsky, and Ray Chwasz of the Chicago Transit Authority, Calvin Francis, Joseph Dorsey, and Mitchell Guidry of the New Orleans Regional Transit Authority, Mike Greenwood, Mark Maloney, Alan Jagger, Vince Greene, Randy Jumper, Dean Delgado, and Frank Lonyai of the Orange County Transportation Authority, TJ Ross of the Phoenix Regional Public Transportation Authority, who provided me with a lot of useful information on the electronic farebox and automatic passenger counter and shared with me their knowledge and experience with it.

I would like to thank Ching Law and my brother Banny Wong who lent me their precious disk space for saving my backups. In addition, many thanks go to the people who lent me (or let me dial up to) their machines when I needed them the most. These people include, but are not limited to, Eugene Fung, Kenneth Hon, Salal Humair (who gave me access to glo-worm), Ally Ip, Warren Lam, Monique Lo, Alex Shing, Joseph Wu, Annie Xia, and Danny Yim. Thanks also go to the people who supported me spiritually (by zephyring me or eating with me) throughout the whole research project, actually over my past 6 years at MIT.

Last but not least, I would like to thank Professor Nigel Wilson for his guidance and support, both intellectually and financially, and not only on this research project, but on my college life as a whole since my freshman year here at MIT. Once again, thank you.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The primary objective of this thesis is to investigate the feasibility of relying primarily or exclusively on Electronic Fareboxes (EF) to collect the information needed for monitoring the performance of public transport services at the individual route level. Currently, most such information is collected by traffic checkers. There are two disadvantages associated with using traffic checkers for data collection. First, this practice is very expensive and labor intensive. Second, it provides an imperfect view of operations due to errors in observations and limitations in the sample size observed due to high data collection costs. Therefore, a more cost-effective technique for data collection is desirable. New technology such as the electronic farebox provides new opportunities for large scale data collection at low cost. Since electronic fareboxes can be installed on all buses, and they only require operators to press the correct button(s) at the start of each trip and as each passenger boards the bus, they have the potential to provide data for all trips at low cost.

1.1 Purpose of Data Collection

In order to use public transportation resources efficiently, it is important to evaluate from time to time all services, both current and planned. For example, if the ridership of a route has decreased, service on this route may be reduced, and some of the resources for this route may be used on another route which has increased ridership. In order to evaluate transit services, performance data such as revenue and ridership are needed. Therefore, a well-designed and cost-effective data collection program is very important for service planning in any transit agency. In addition to service planning, data collection is important for reporting purposes. Some outside agencies such as local and state government, and the Federal Transit Administration may require transit agencies to report to them periodically specific systemwide passenger-related data.[1]

1.1.1 Data Needs

Transit agencies need many different kinds of data to support their operations. This includes qualitative data such as passenger attitudes and travel needs, useful for market research and service planning, but which can only be gathered through customer surveys. On the other hand, there are quantitative data such as passenger usage and vehicle movements, which can be obtained through some form of observation of the system. These data are useful for monitoring the performance of individual services and for input into the scheduling and operations planning functions. Furthermore, some of these data are required to be reported to the federal government at the system level.

This thesis focuses on the collection of passenger usage data because the use of electronic fareboxes has the potential to improve the collection of this kind of data. Specifically, the collection of three types of passenger usage data (daily revenue by route, daily ridership by route, and maximum passenger load by route and time period) may be improved by the proposed use of electronic fareboxes. Daily revenue and ridership by route are key inputs to overall route level performance with respect to several measures. Daily fare revenue and ridership by route can be generated directly by the electronic farebox system using appropriate software. Unfortunately, driver entry errors and farebox deficiencies often cause these data to be unreliable. Nevertheless, these raw data provide a

1. United States Department of Transportation, "*Transit Data Collection Design Manual*", June 1985.

basis for estimation of the target information. Daily pass revenue by route must be estimated from the number of swipes by passes of different kinds at the route level.

Another type of passenger usage data of interest is the maximum passenger load by route and time period. This kind of data are used for frequency determination and vehicle scheduling, and may be estimated by a conversion factor based on total ridership by route and time period.

Since the estimation of pass revenue from the number of swipes by passes of different kinds depends on conversion factors independent of the electronic farebox mechanism; in order to eliminate possible errors introduced by these conversion factors (rather than the electronic farebox mechanism), only the daily number of pass swipes by different kinds of passes by route are studied in this thesis. Similarly, since the estimation of maximum passenger load by route and time period from ridership by route and time period depends on conversion factors independent of the electronic farebox mechanism, only ridership by route and time period is studied in this thesis to eliminate possible errors introduced by the conversion factors. Therefore, this thesis focuses on the following four types of passenger usage data:

1. Daily Fare Revenue by Route
2. Daily Pass Swipes by Route (for estimating Daily Pass Revenue by Route)
3. Daily Ridership by Route
4. Ridership by Route and Time Period (for estimating Maximum Passenger Load by Route and Time Period)

1.1.2 The Three Phases of Data Collection

In general data collection activities can be divided into three phases: baseline, monitoring, and follow-up. The first phase is the baseline phase in which all data needed for service evaluation and planning are collected to define the base conditions for each route in the system by time period. One important use of the baseline data is to estimate relationships among data items which may be used to reduce the effort needed to monitor performance in the future. These relationships are called conversion factors. For example, if the maximum passenger load for a particular route is closely correlated with the total ridership for that route, and if total ridership data are less costly to acquire than maximum passenger load data, then ridership can be used to estimate the maximum passenger load for that route to avoid collecting maximum passenger load data directly. The baseline phase is traditionally conducted by manual techniques such as the ride check although some properties have now begun to use Automatic Passenger Counters for this purpose. Since this phase is expensive to conduct, it is typically done infrequently, perhaps once every 5 - 10 years.

The monitoring phase is conducted periodically for each route in the system and generally consists of gathering a more limited set of data than in the baseline phase. This phase has two roles. The first involves the tracking of data items that need to be updated often because they are very important to the regular scheduling process or are necessary to meet reporting requirements. The second involves checking some key data items (change indicators) to decide whether a significant change has occurred which would trigger a follow-up data collection activity. If none of the change indicators suggest a significant change, then it is assumed that the conversion relationships identified in the baseline phase still hold, and a follow-up procedure is not necessary. This phase is conducted relatively frequently (up to four times per year) on every route in the system, and it is the most expensive element of most data collection programs. It is also the phase that would be affected by the proposed use of electronic fareboxes for data collection.

The final phase is called the follow-up phase. This phase has two purposes. The first is to provide more accurate and current data than those provided by the baseline and monitoring phases required for some planning applications. The second purpose is to update the data gathered in the baseline phase when a major change is detected in the monitoring phase or when there is a major external change such as a fare increase. In general, only the affected routes are updated, and the results become the baseline conditions for future monitoring phases.

1.2 Traditional Data Collection Methods

There are three major traditional data collection methods: the ride check, the point check, and the driver check.

1.2.1 The Ride Check

In a ride check, a traffic checker rides on the bus as it travels along the route. This deployment option can be used to gather various kinds of information such as load, ridership, running time, and revenue. In addition, ride checkers can also distribute surveys to passengers during a trip to gather further passenger information. However, the ride check is a relatively expensive deployment option because it consumes one checker-hour for every vehicle-hour of service monitored. The ride check is the principal method of manual data collection in the baseline phase.

1.2.2 The Point Check

In a point check, a traffic checker is stationed at roadside and observes buses as they pass by, although in some cases the traffic checker briefly boards the bus to get a more accurate passenger count (this is particularly useful given the prevalence of tinted windows). Although point checks can still provide a variety of information, the information is less comprehensive than that gathered by ride checks. However, point checks are generally less expensive than ride checks because point checks only consume one checker-hour per point per hour of observation in each direction. In addition, there are ways to reduce the cost further. For example, if a traffic checker can observe traffic in both directions, the cost is halved.

There are four different types of point checks. The first type is the peak point check in which a traffic checker is stationed at the roadside at the point on a route where the average passenger load is the greatest. Since the peak point changes from time to time, it has to be verified periodically by a ride check. The second type is the multiple point check in which traffic checkers are stationed at several points along a route instead of just a single point. This option can be useful on long routes or routes that serve a number of important activity centers because such routes may have a number of points at which information is important for planning purposes. The third type is the endpoint check in which a checker is stationed at each end of a route. This option is very useful for monitoring run time and trip level revenue information if the buses are equipped with registering fareboxes. If only one checker is used instead of two, only roundtrip information can be monitored. If an endpoint is a terminal for many different routes, such as a rapid transit station, then endpoint checks can be very efficient. The fourth type is the undesignated point check in which a traffic checker is stationed at a selected point which is not necessarily the peak point. If the selected point is a point where many routes pass, and the data item being monitored need not be recorded at the peak point, then the undesignated point check is a very economical way of monitoring information on these routes.

1.2.3 The Driver Check

In a driver check, the driver records some data as he or she operates the bus. Although this method is very inexpensive because traffic checkers are not needed in this deployment option, it has two disadvantages. First, work rules and union contracts can severely limit the amount of data collection work that can be assigned to drivers. Second, since most drivers consider data collection at best a secondary duty, many are not fully committed to it, and the collected data are generally less reliable than those collected by traffic checkers. The issue of driver participation and commitment is at the heart of the research in this thesis, and so this topic will be addressed at greater length later.

1.3 New Technology for Data Collection

Four new tools have been developed to complement the traditional methods for data collection:

1. Electronic farebox (EF)
2. Automatic passenger counter (APC)
3. Automatic vehicle identification (AVI)
4. Automatic vehicle location (AVL)

Although the sole, or even the primary, purpose of each of these systems is not data collection (for example, the primary purpose of the EF system is to facilitate revenue capture, and the primary purpose of the AVI and AVL systems is to improve real-time operations control) the above systems can be installed individually or in combination to improve the data collection process in transit agencies. This thesis will focus on the value of the electronic farebox in improving data collection.

1.3.1 Electronic Farebox (EF)

The electronic farebox is a fare-collection device which can perform the functions of a traditional farebox such as fare collection, fare storage, and fare count, plus a number of other new features such as pass reading and verification and ridership counting. Since the research in this thesis focuses on the electronic farebox, a detailed description of the electronic farebox is given in section 1.4.

1.3.2 Automatic Passenger Counter (APC)

Automatic Passenger Counters record the number of passengers boarding and alighting at each stop along with auxiliary information such as time or distance measurements, or coded location signals transmitted to the bus from devices mounted on signposts, to facilitate matching boarding and alighting information with bus stops. The APC acquires data by sensors that are located at each doorway of the bus to detect passenger movement. The data collected is then recorded and stored temporarily in a data processing unit. The data is transferred to a central computing facility periodically where appropriate software is used for analysis and to generate the desired reports. Typically only 10% to 15% of buses in a fleet have to be equipped with APC.[1]

There are a number of advantages associated with the APC. First, although the initial investment is substantial, the incremental cost of its usage is low. Depending on the capabilities, features, and the number of buses equipped with APC, an APC system typically costs between $1,500 and $15,000 per bus.[2] The cost per bus of APC generally decreases with increasing number of buses which are APC-equipped. Second, the APC

---

1. United States Department of Transportation, Federal Transit Administration, "*Advanced Public Transportation Systems: The State of the Art Update '96*", January 1996.
2. *ibid.*

can collect many different types of data such as seasonal or day-to-day variations in ridership at relatively low cost compared with manual checks such as ride checks or point checks. Third, the APC can generate reports at different levels of detail within a reasonably short amount of time, which is hard to achieve from manual checks. Therefore, the APC is attractive to large transit systems which have substantial data needs, but are not satisfied with their manually collected data.

1.3.3 Automatic Vehicle Identification (AVI)[1]

Automatic vehicle identification systems are generally used to provide transit supervisors and controllers information about the location of the vehicles. Each vehicle is equipped with one or more encoder and transponder. The encoder generates the vehicle identification signal, which is then transmitted to the wayside receiver by the transponder. The vehicle number, the route number, and some other information can be transmitted by the transponder. To give one example, on an MBTA Green Line train, the operator has to enter the route number of the train, and each transponder can transmit 19 binary bits of information. The wayside receiver decodes the signal, sets the track switches appropriately for the rail vehicle, and then passes the information to the control center. Therefore, although the control center does not know exactly where a train is, it knows which track segment the train is in.

Although AVI systems are used primarily for track switching and real-time operations control, they can also be valuable in terms of collecting running time and service reliability data for performance monitoring and planning and scheduling. With the appropriate software, AVI systems can easily produce running time and schedule

---

1. Robert Fellows, "*Using and Enhancing Automatic Vehicle Identification to Improve Service Control on the MBTA Green Line*", Master of Science in Transportation Thesis, Massachusetts Institute of Technology, September 1990.

adherence data at different levels of detail. Like the APC, the initial cost of installing an AVI system is substantial, but the incremental cost for its subsequent use is low.

1.3.4 Automatic Vehicle Location (AVL)[1]

Automatic Vehicle Location systems can generally provide transit agencies with transit vehicle location information, public information, service control information, and management information. There are four types of location systems commonly used by transit agencies. The first three types have been used by transit agencies for a number of years whereas the fourth type is relatively new.

The first is the radio location (pulse trilateration) system in which each bus detects the time differences between radio signals sent from different transmitters by means of a special radio receiver. Since the locations of the towers are known, the location of the buses can easily be determined by measuring the time differences between the radio signals. The regular bus radio then transmits the location data to the control center.

The second is the dead reckoning system in which the bus is equipped with distance and direction sensors to determine the movement of the vehicle relative to the known origin. Distance measurements are given by a high precision odometer on the bus and the direction changes are detected when the bus is steered. The bus radio then transmits the location data to the control center. The actual vehicle path has to be reset periodically at known locations to reduce the magnitude of accumulated errors of the sensors.

The third is the signpost and odometer system in which the location of the bus is determined by the relationship between the bus and fixed low-strength radio transmitters (signposts) placed along the bus route. When the bus approaches a signpost, it receives the

1. Canadian Urban Transit Association, "*Automatic Vehicle Location and Control Workshop Proceedings*". November 15, 1987.

signals generated by the post and sends its locational data to the control center. Generally, the odometer is used to estimate the vehicle's location between signposts.

The last is the Global Positioning System (GPS) which is currently the most commonly used technology for new systems. GPS technology uses signals transmitted from a network of 24 satellites in orbit. Receivers are placed on the roof of each bus. Signals from the satellites are received by each bus, and the location of each bus is transmitted to dispatch. There are two advantages associated with GPS. First, since the signals are transmitted by satellites, the coverage is very broad. Second, the incremental cost is low because only receivers are needed since the satellites are already in orbit. The only shortcoming of GPS is that the signals cannot be received by the bus when it is underground or near tall buildings. Hence GPS is often supplemented by one of the above three technologies.[1]

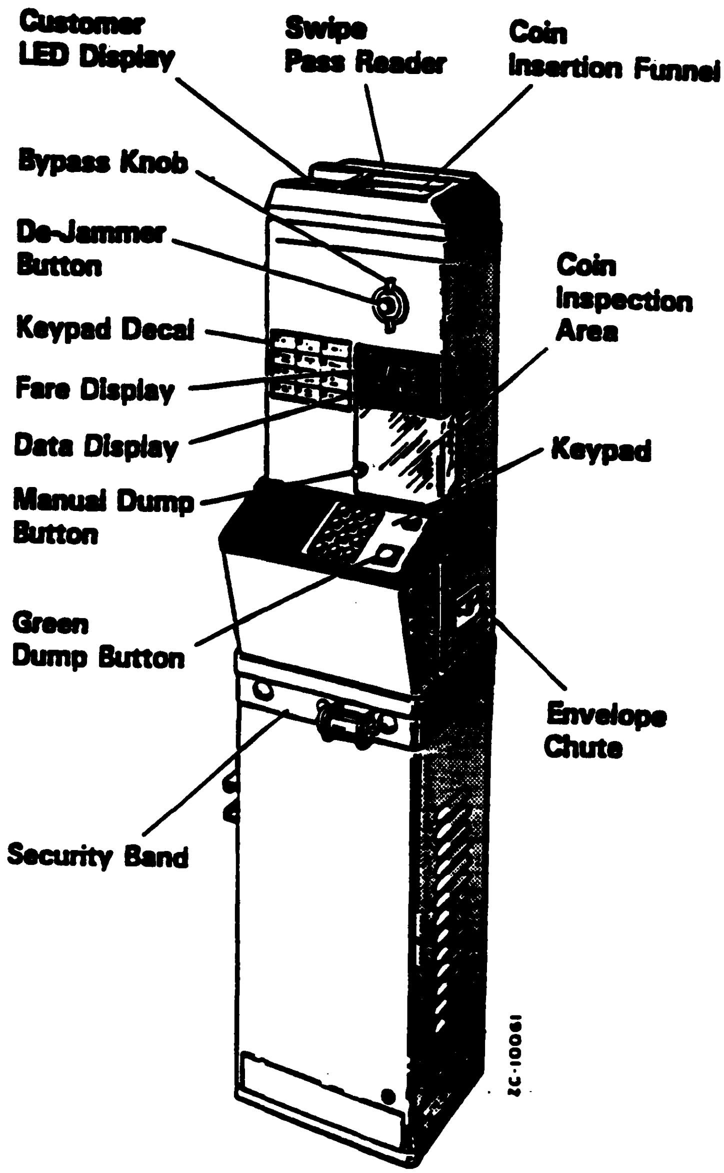
Although the AVL system is primarily used for real-time operations control and security enhancement in an emergency such as an accident, a fire, or a passenger disturbance, there is obvious potential for the data generated by the AVL system to be used for service planning, management, and reporting purposes. With the appropriate computer software, the AVL system can generate information such as schedule adherence, running time, passenger load, and vehicle mileage at different levels of detail.

1.4 The Features of the Electronic Farebox[2]

This section gives a detailed description of the electronic farebox which is shown from the driver's view in figure 1.1.

1. United States Department of Transportation, Federal Transit Administration, "*Advanced Public Transportation Systems: The State of the Art Update '96*", January 1996.
2. Massachusetts Bay Transportation Authority, "*Farebox Repair, Maintenance, & Service Manual M00511-3SM*".

Figure 1.1: The Driver View of an Electronic Farebox[1]



1. Massachusetts Bay Transportation Authority, "*Farebox Repair, Maintenance, & Service Manual M00511-3SM*", figure 1-1.

1.4.1 Farebox Design Objectives

The electronic farebox is designed to meet a number of different objectives. First of all, the electronic farebox accepts all valid coins, tokens, and passes. Second, the farebox helps operators determine if a passenger has paid the correct fare by providing a "beep" when a correct full fare is paid or a valid pass is read. Third, the farebox facilitates data collection by registering an accurate total of all the inserted fares, storing revenue and ridership data, and having the capability of generating reports on collected data. Finally, the farebox ensures the security of revenue and the revenue transfer process.

1.4.2 General Operation of the Electronic Farebox

The farebox is installed on the vehicle with the panel facing the operator so that the operator can see the displays and press the keys as appropriate. The boarding passengers also sees a display of the fare paid so that they can pay the appropriate fare by dropping the coins or tokens into the coin insertion funnel on the farebox or swiping the pass through the swipe pass reader. Before the start of each trip, the operator has to key in the fareset of the trip, the operator's badge number, the signcode of the trip, and the run number so that the farebox knows what the standard full fare should be and which trip (and route) the revenue and ridership belong to.

When a passenger boards the bus, the inserted coins are counted as they enter the farebox, and the value is shown on the display and is recorded. In addition, tokens and passes are also recorded. If a passenger pays a full fare, the farebox will "beep" and reset to zero to show that it is ready for the next fare. This way, the operator will know if the passenger has paid the correct fare and if not, how much money the passenger paid. If a full fare is not inserted, the operator has to identify the special fare (such as student, senior citizen, or disability fare) and press the appropriate button on the key pad so that the inserted money will be recorded correctly. In addition, there is a green dump button which should be used when a passenger overpays. This button drops the coins into the vault and resets the operator's display to zero to show that it is ready for the next fare. Finally, at the end of each trip, the operator has to press an appropriate sequence of keys to make a trip record for the completed trip. Figure 1.2 shows the fare keys on an electronic farebox.

In addition to the special fare buttons, there are three exception buttons that are used when a special situation occurs. The manual dump button allows the operator to dump coins from the escrow to the cashbox and is used when there is a coin jam or power failure. The de-jammer button loosens jams in the coin mechanism and is used when a jam occurs. Finally, the bypass knob is used when the coin mechanism is badly jammed or defective. This knob turns the fare display off and in this case, the electronic farebox will not count any inserted money, and it becomes a non-registering farebox. All three exception buttons affect the quality of the trip record because the fares are not counted correctly when any one of them is pressed.

1.4.3 Data Collection for Service Planning by the Electronic Farebox

Although many transit agencies have installed electronic fareboxes primarily to facilitate revenue capture, the electronic farebox can record revenue and ridership data which are very useful for service planning purposes. Some data are printed in daily reports by the vendor's computer software, and the more detailed data are stored on a mainframe computer.[1] Revenue and ridership data are available at different levels of detail such as by

1. James Mulqueeny Jr., Sarah Labelle, Ross T. Patronsky, Joseph Simonetti, "*What to Do With Your New Electronic Farebox Data*", Chicago Transit Authority, Presented at TRB Conference on New Public Transportation Technology, San Francisco, August 1992.

Figure 1.2: Bus Fare Keys on an Electronic Farebox[1]

1 Student/Child	2 Senior Citizen	3 Disability
4 Student Pass	5 Visitor/10-Ride	6 Free
7 Pass Accepted/ Transfer Received (Special routes only)	8 Other Full Cash/ Full Transfer Issued (Special routes only)	9 Other 1/2 Fare Reduced Transfer Issued (Special routes only)
* Short Fare	0 Hold/Display	# Enter

1. Massachusetts Bay Transportation Authority, "*Farebox Repair, Maintenance, & Service Manual M00511-3SM*", page 1-4.

garage and day, by signcode and day, and by signcode and trip; in addition, different fare types and pass types are also recorded. Furthermore, the electronic farebox data may have the potential to be used for estimating maximum passenger load, using a conversion factor based on ridership.

However, there are two issues that need to be recognized. First, the electronic farebox requires a good level of driver compliance to operate satisfactorily. If drivers do not press the correct buttons, data will be inaccurate. In addition, some revenue and ridership will be unclassified or unknown because the electronic farebox cannot associate that revenue and ridership with a particular trip or route. In order to solve this problem, some transit agencies use a farebox adjustment factor to account for the unclassified or unknown data. That is, the electronic farebox revenue and ridership data are factored up by an adjustment factor to better estimate the true revenue and ridership.

Second, some computer software should be developed in-house. This software is needed to obtain revenue and ridership information from the electronic fareboxes, to validate the data, to store the information for future use, and to provide operators with feedback on their mistakes. Although the farebox vendors can develop this kind of software, it tends to be focused on revenue capture rather than service planning; hence it is better for this kind of software to be developed by the transit agencies themselves.[1]

1.5 The Purpose of the Research

The primary objective of the research is to investigate the feasibility of relying primarily or exclusively on electronic farebox technology to collect the information

1. Nigel Wilson, Peter Furth, MacDorman & Associates, "*Review of MBTA Data Collection Plan*", Design of Service Quality Measures and Planning Standards, prepared for the Massachusetts Bay Transportation Authority by the Conservation Law Foundation, Inc., July 11, 1995.

needed for monitoring the utilization of public transport services at the individual route level. As mentioned above, current electronic farebox data are imperfect because of driver compliance problems as well as limitations of the fareboxes themselves. The research will explore ways of estimating different forms of utilization information based on actual (and hence imperfect) electronic farebox data by minimizing the effects of problems such as poor driver compliance and defective fareboxes. The research will focus on four types of passenger usage data:

1. Daily Fare Revenue by Route
2. Daily Pass Swipes by Route (for estimating Daily Pass Revenue by Route)
3. Daily Ridership by Route
4. Ridership by Route and Time Period (for estimating Maximum Passenger Load by Route and Time Period)

The reason for focusing on the four types of data above is that the number of swipes by different kinds of passes by route is necessary for estimating pass revenue by route and day; and together with daily fare revenue by route, there is potential that the total daily revenue by route can be estimated. In addition, ridership by route and time period, with the appropriate conversion factor, may be used to estimate maximum passenger load by route and time period.

The result of the research will be important to medium and large transit agencies interested in taking full advantage of existing electronic farebox technology because many of these transit agencies still rely on traditional methods for gathering the data necessary for good operations planning. This means that they are still relying on manual traffic checkers to conduct a combination of ride checks and point checks to gather data needed for the scheduling and operations planning functions.[1] These are expensive and labor-

---

1. Transportation Research Board National Research Council, "*Transit Cooperative Research Program Synthesis 10*", page 44, sponsored by the Federal Transit Administration, 1995.

intensive activities which provide only an imperfect view of operations due to errors in observations and limitations in the sample size due to high data collection costs.

Although some transit agencies are beginning to use electronic fareboxes for data collection, many are still not satisfied with data quality. This research hopes to provide transit agencies with ways of estimating useful information such as revenue and ridership from raw electronic farebox data. In other words, the research hopes to establish ways of eliminating or reducing the effects of problems such as poor driver compliance or defective fareboxes on electronic farebox data. If this proves possible it would reduce the cost and labor-intensiveness of data collection for many transit agencies, and valuable resources could be reassigned to other areas that are desperately in need.

1.6 Research Approach

The research focuses on the experience of the Massachusetts Bay Transportation Authority (MBTA) with electronic fareboxes. A set of methods are proposed to estimate the data items of interest and these methods are then applied to the electronic farebox data available at the MBTA. Data from two years are analyzed separately, winter and spring 1994, and winter and spring 1996. The 1994 data was selected because it was in the early days of EF technology usage at the MBTA, and driver compliance was spotty. In 1996 the MBTA was beginning a serious initiative to use EF data for route performance monitoring, and a major effort was being made to increase driver compliance and correct usage of the farebox. Thus the results for 1996 might be more representative of what is achievable in other transit agencies with more established EF technology.

1.7 The Massachusetts Bay Transportation Authority (MBTA)[1]

Since this thesis is based heavily on the experience of the Massachusetts Bay Transportation Authority (MBTA), it is appropriate to provide some background on the authority. The MBTA is a multimodal transit agency in the eastern part of Massachusetts which serves a population of more than 2.6 million in 78 cities and towns covering an area of 1,038 square miles. It includes 3 heavy rail lines (Red, Orange, and Blue Lines), 1 light rail system with 5 branches (B, C, D, E Green Lines, and the Mattapan High Speed Line), 4 trackless trolley lines, 11 commuter rail routes operated under contract by Amtrak, more than 150 bus routes, and some other services under contract with other companies such as a handicapped service (The Ride), a commuter boat service, and some suburban bus lines. The MBTA owns 1,107 buses, 219 light rail vehicles, 12 PCC's, 50 trackless trolleys, 52 commuter rail locomotives, 304 commuter rail coaches, 129 specially equipped vans (an additional 120 specially equipped vans are contractor-supplied). The average weekday ridership is 922,000 and the commuter rail carries 74,000. The bus routes are operated out of 9 garages: Albany, Bartlett (Arborway), Cabot, Charlestown, Fellsway, Lynn, North Cambridge, Quincy, and Somerville.

1.8 The MBTA Experience with Electronic Fareboxes

The MBTA introduced electronic fareboxes in the summer of 1993 and has been interested in using the technology for improving data collection since then. However, the agency still relies heavily on manual checks (both ride checks and point checks) for data collection. Four aspects of the MBTA experience with electronic fareboxes were reviewed:

1. Massachusetts Bay Transportation Authority (MBTA) Homepage: http://www.mbta.com

1. Driver compliance, driver training, and farebox problems
2. The use for planning and route performance evaluation
3. The transit agency's expectations
4. The validation of electronic farebox data

1.8.1 Driver Compliance, Driver Training, and Farebox Problems[1]

The MBTA uses GFI fareboxes, which record revenue and ridership data including special fares and passes, and produces reports at different levels of detail such as by route and day, by route and trip, etc. The electronic fareboxes generally undercount ridership by about 10%, but no information on revenue undercount is available. The MBTA believes that there is room for improvement in driver compliance, and has taken many measures to encourage driver compliance. First of all, extra efforts were made to remind drivers of the proper farebox entry procedure. For example, a quick reference guide is given to each driver, with different versions available for buses, trackless trolleys, and streetcars. It lists all keypad labels and tells drivers how to classify the different types of passengers and press the right keys on the electronic farebox. It also reminds the drivers to enter the signcode after each half roundtrip, reviews the proper log on, signcode, and log out procedures, and discusses the fare exceptions such as transfers. Besides, a special order was posted stressing the need for compliance with electronic farebox data entry guidelines. In addition, stickers were placed on the fareboxes reminding operators to enter new signcodes after each half-trip. Furthermore, the electronic destination sign and farebox code listings were made readily available to all operators.

Second, training and retraining programs are available for drivers. Each training or retraining session on the electronic farebox lasts about an hour. About half of the drivers

1. Alan Castaline and Clinton Bench, "*Memorandum on Operator Compliance with Electronic Farebox Data Entry Guidelines*", April 18, 1996.

in Bartlett Garage were returned for retraining, and other garages may soon follow suit. The preliminary results of the retraining program were very promising at both the route and garage levels. Based on five selected routes at Bartlett Garage, the number of trip records and passengers recorded increased dramatically, as did the number of different kinds of special fare passengers such as students, children, and senior citizens recorded. At the same time, the amount of unclassified revenue as a portion of total revenue decreased, as did the number of times operators pressed the "dump" button.

Third, although no incentives are being provided and no disciplinary action taken, a special order was given requiring each driver to comply with the farebox entry procedures and warning that disciplinary actions may be taken for offenders. However, no disciplinary action has yet been taken for any driver even though several of the worst offenders have been called in by the garage managers to discuss their performance.

Finally, the MBTA tries to identify the problem operators, currently, focusing on Bartlett Garage. The performance of the drivers on one heavy route (Route 39: Copley Station - Forest Hills) is closely monitored. In addition, computer software is used to detect log on problems and excessive use of the dump buttons on all Bartlett Garage routes. Furthermore, systemwide Sunday data are carefully studied to detect non-compliance.

1.8.2 Use for Planning and Route Performance Evaluation

The MBTA is intending to use electronic farebox data for planning purposes in the form of daily revenue by route, and ridership by route and trip. In addition, maximum passenger load information for Route 1 (Dudley Station - Harvard Station) has been estimated from electronic farebox ridership data on a trial basis. Ridership information will be used for route level service analysis. Since ride checks are not routinely performed on Saturdays, the MBTA is using electronic farebox ridership data for Saturdays to estimate the demand on Sundays, based on pre-established conversion factors relating Sunday ridership to Saturday ridership, for routes that are being considered for additional Sunday service.

1.8.3 MBTA Expectations

The MBTA is expecting better driver compliance in the future as a result of the several measures taken to encourage driver compliance and more convenient data queries as a result of software improvement. However, due to uncertainty in personnel and budget, the MBTA does not know when these expectations will be realized. In addition, the MBTA is hoping to begin performing annual service evaluations using electronic farebox ridership data in about 2 years.

1.8.4 Validation of Electronic Farebox Data

The MBTA currently does not validate electronic farebox data systematically or rigorously. Occasionally electronic farebox data are compared with available ride check data at the route and trip level. Generally, an average weekday ridership figure (by route and day) is calculated based on ride checks and is compared with 4 times the Sunday electronic farebox ridership figure (by route and day). If the two figures are similar, then the electronic farebox data are considered good because experience shows that Sunday ridership is roughly 25% of that of a typical weekday. On the other hand, validation is not usually done on electronic farebox revenue data unless the revenue department notices major problems with the farebox data. The revenue department generally uses the electronic farebox revenue data directly without any modification.

1.8.5 Summary of the Experience of the MBTA on the Electronic Farebox

From above, we can see that the MBTA is beginning to use electronic farebox data for planning purposes even though it is still relying heavily on manual checks. In addition, the MBTA believes that driver non-compliance is the major obstacle to the successful use of the electronic farebox for service planning, and has taken a number of measures to increase driver compliance. It expects increased driver compliance as a result of the measures taken and more easily accessible data as a result of software improvements in the future.

1.9 Prior Research

A significant amount of research has been done on the role of new technology in data collection for public transportation, but most has focused on the technical aspects of the tools without much emphasis on the use of the collected data for service planning and monitoring purposes. In the *Canadian Urban Transit Association Workshop Proceedings on Automatic Vehicle Location (AVL) and Control (1987)*[1], most of the effort is dedicated to the technical aspects of the AVL system and its potential functions, but not much is said about how to use the collected data of the AVL system for service planning and monitoring purposes. In addition, in *Advanced Public Transportation Systems: The State of the Art Update '96*[2], the focus of the electronic farebox discussion is on fare collection, not data collection. Similarly, in *Integrating Electronic Fareboxes at the CTA*[3], the main focus is on the hardware and the software of the electronic farebox system of CTA. There

1. Canadian Urban Transit Association, "*Automatic Vehicle Location and Control Workshop Proceedings*". November 15, 1987.
2. United States Department of Transportation, Federal Transit Administration, "*Advanced Public Transportation Systems: The State of the Art Update '96*", January 1996.
3. Thomas C. Dawson, Dennis Ryan, "*Integrating Electronic Fareboxes at the CTA*", Chicago Transit Authority, December 1993.

is not much discussion on the use of the collected electronic farebox data for service planning and monitoring purposes.

Although some research has been done on using the electronic farebox data for service planning and monitoring purposes, most of the methods described simply factor up the recorded data uniformly across the system to account for the unclassified and unknown data. For example, in W*hat to Do With Your New Electronic Farebox Data*[1], the Farebox Adjustment Factor and the Ridership Adjustment Factor are used uniformly across all routes to account for the unclassified or unknown data. While the heavy reliance on expensive manual traffic checkers[2] and the poor quality of raw electronic farebox data are recognized[3], there is no work on minimizing or eliminating the effects of problems such as poor driver compliance or defective fareboxes. In this thesis, emphasis will be placed on exploring different ways of estimating the true revenue and ridership information based on electronic farebox data.

1.10 Thesis Contents

Chapter 2 discusses the experience of other transit agencies with electronic fareboxes based on telephone interviews with eight transit agencies across the United States. Chapter 3 describes the proposed methods for using the electronic farebox data to better estimate the true passenger usage information. Again, the focus is on daily fare revenue, pass swipes, ridership by route, and ridership by route and time period. Chapter 4 discusses the

---

1. James Mulqueeny Jr., Sarah Labelle, Ross T. Patronsky, Joseph Simonetti, "*What to Do With Your New Electronic Farebox Data*", Chicago Transit Authority, Presented at TRB Conference on New Public Transportation Technology, San Francisco, August 1992.
2. Transportation Research Board National Research Council, "*Transit Cooperative Research Program Synthesis 10*", page 44, sponsored by the Federal Transit Administration, 1995.
3. Nigel Wilson, Peter Furth, MacDorman & Associates, "*Review of MBTA Data Collection Plan*", Design of Service Quality Measures and Planning Standards, prepared for the Massachusetts Bay Transportation Authority by the Conservation Law Foundation, Inc., July 11, 1995.

tests used to determine the validity of the methods and presents the results of the analysis and the conclusions. Data from the winters and the springs of 1994 and 1996 are used to study the effect of the change in driver compliance. Chapter 5 summarizes the project findings and discusses potential further research directions.

Chapter 2

The Experience of Transit Agencies in the United States

This chapter discusses the experience of other transit agencies in the United States with electronic farebox technology. The purpose is both to place the MBTA experience in a broader context and to ensure that any ideas which have proven helpful in other agencies are identified for the MBTA phase of the research.

2.1 Introduction

In order to understand the experience of other transit agencies with electronic farebox technology, a number of telephone interviews were conducted. A total of eight transit agencies across the nation were included in this review, as listed below in descending order of the number of buses in the fleet[1]:

1. New York City Transit Authority (NYCTA)
2. Chicago Transit Authority (CTA)
3. Washington Metropolitan Area Transit Authority (WMATA)
4. Seattle Metro Transit Authority (Seattle Metro)
5. Greater Cleveland Regional Transit Authority (Cleveland RTA)
6. New Orleans Regional Transit Authority (New Orleans RTA)
7. Phoenix Regional Public Transportation Authority (Phoenix RPTA)
8. Orange County Transportation Authority (OCTA)

These agencies were selected to include those with extensive experience with EF technology and those with demonstrated interest in making full use of the data being produced through the fareboxes. All agencies selected use GFI fareboxes except NYCTA and Phoenix RPTA, which uses Cubic fareboxes (in order to facilitate integration with the

1. Federal Transit Administration (FTA) Homepage: http://www.fta.dot.gov

subway system) and Duncan Faretronix fareboxes, respectively. As with the MBTA assessment presented in chapter 1, four major areas were studied for each agency:

1. Driver compliance, driver training, and farebox problems
2. Use of EF data for planning and route performance assessment
3. Agency expectations
4. Validation of electronic farebox data

2.2 History of EF Use

In this section the history of EF technology used in each agency is briefly described.

The Office of Management and Budget of NYCTA has been using electronic fareboxes to track revenue by bus route since 1990, and the Operations Planning Department has also been using this information for planning purposes. Since 1992, NYCTA has been using the fareboxes for senior citizen ridership counts, and since the switch from GFI to Cubic fareboxes in 1995, the use of other ridership counts from the fareboxes is being investigated. NYCTA hopes to be able to use the student ridership counts once students begin using electronic farecards in September 1996. By January 1996, 33% of the subway stations were equipped with the Cubic turnstiles, and all buses were equipped with the Cubic fareboxes.

The CTA switched from Duncan fareboxes to GFI fareboxes about 11 years ago and was one of the first customers for GFI fareboxes. Since then, CTA has improved the vendor's software of the fareboxes, written its own software, and added the magnetic card reader option. The CTA is currently using GFI version 4 software.[1]

---

1. Thomas C. Dawson, Dennis Ryan, "*Integrating Electronic Fareboxes at the CTA*", Chicago Transit Authority, December 1993.

WMATA introduced the GFI fareboxes in 1987 to facilitate revenue capture, but did not use the electronic farebox data for planning purposes until 1992. The GFI fareboxes gather trip level information, send it to a mainframe computer overnight, and then the information is downloaded monthly to PCs. However, the computer system is near the end of its useful life, and there is discussion of getting a new system in the future.

Seattle Metro has upgraded to Version 6.35 for Network Manager and Version 6.33 for Data Manager. The reason for the upgrade is that about 40 different transit passes are now available to Seattle commuters, and the GFI software needs to be upgraded to recognize each in this myriad of payment options. Seattle Metro is also using Paradox database software to manage old farebox data and prepare it for further analysis. In particular, GFI-generated comma-separated variable files are translated into text files which are read by the database program. Individual tables of data are then constructed for each garage.

Cleveland RTA also uses EF technology for both revenue capture and bus route performance evaluation, and it currently uses GFI Software Version 4.60.

Phoenix RPTA Valley Metro's fareboxes were purchased in the early 1980's from Duncan Faretronix. They were recently rebuilt in-house to better meet current data collection needs -- in particular, magnetic-stripe card readers were installed for Valley Metro's employer credit card program. In this program, individual employers may purchase bus credit cards for their employees who use transit for work-related trips. Then at the end of each month, participating employers receive an invoice for these trips -- they may choose to request partial reimbursement from their employees, or cover the full cost themselves.

2.3 Driver Compliance, Driver Training, and Farebox Problems

In general driver compliance is a problem at all transit agencies, although the magnitude of the problem varies. Chicago, New Orleans, Phoenix, and Orange County experience only minor driver compliance problems, but Washington DC, Cleveland, and Seattle experience more significant driver compliance problems. New York City experiences some driver compliance problems although the severity is unknown. Driver non-compliance leads to undercounts of revenue and ridership in almost all agencies, however, and must be recognized.

2.3.1 New York City Transit Authority (NYCTA)

NYCTA knows that there are some driver compliance problems, but it is not clear how severe they are. Although NYCTA does not have quantitative figures showing the size of the problem, the agency is aware of the problem through missing data for some trips. However, since not many bus routes are interlined in NYCTA, the daily data are thought to be reasonably reliable. The Office of Management and Budget has determined that 97 to 98 percent of the revenue is collected on trips in which the bus has a "valid" destination code, i.e. a destination code for a route that runs out of the assigned depot for that bus. However, this does not necessarily mean that the destination code is correct, or that the operator is remembering to enter a new code for each trip. It also says nothing about the accuracy of the ridership counts.

NYCTA carries out a number of actions to improve driver compliance. First, disciplinary actions can be imposed on problem operators through the depot managers, although the details are not known. Second, a 4-hour training session and a comprehensive manual on the electronic farebox are provided for each driver. Third, a retraining session will be scheduled some time in 1996. Fourth, for the newer buses (purchased in or after 1990), the destination signs are integrated with the electronic fareboxes through communication links. Thus, drivers need only enter information through the electronic farebox at the beginning of each trip. Finally, the software system tracks driver numbers in order to identify problem drivers. However, since NYCTA is still at the early stage of using electronic fareboxes for planning purposes, no report on the resulting improvement is yet available.

2.3.2 Chicago Transit Authority (CTA)

CTA says that 92% of cash and tokens, 91% of pass usage, 85% of transfers, and 31% of free riders (parking enforcement employees, CTA employees, children under 7) are recorded properly. The most common problems are improper or missed registration of fare types, miscounts of ($1.25) tokens as dimes, users flashing their passes instead of swiping them through the pass reader, and the normal wear-and-tear on the electronic fareboxes. The agency also feels that since it is large, occasional driver entry errors should not have a large effect on data validity.

Since CTA has a relatively high driver compliance rate, it is not aggressively trying to improve driver compliance. For example, although the CTA could detect problem operators, it is not currently doing so. In addition, no incentives or disciplinary actions are taken for drivers based on their performance in operating the electronic farebox. While training is provided for drivers, retraining is provided only when there is a major change in the keys of the electronic farebox.

The agency routinely adjusts electronic farebox data to improve the revenue and ridership estimates. For example, in order to account for farebox undercounts and miscounts, the agency has developed two factors: the farebox adjustment factor (FAF) and the ridership adjustment factor (RAF). The FAF accounts for the improper registration of fare types by operators and the miscounts of tokens as dimes. The FAF is developed from GFI farebox ridership data and ridership data observed by ride checkers. The RAF accounts for ridership outside the GFI system such as rear door boarding. It also accounts for other farebox errors that are not covered by the FAF. The RAF is developed from GFI farebox revenue data and General Ledger revenue reports. Both RAF and FAF are computed at the system level.

2.3.3 Washington Metropolitan Area Transit Authority (WMATA)

WMATA realizes that driver compliance is a problem, but the extent of this problem has not yet been quantified on a system-wide level. The agency is aware of the problem through missing data for some trips and the generation of non-zero revenue on some non-revenue trips. Since some bus routes are interlined, sometimes information for a particular bus route is attributed to another route when the operator does not register properly before each trip. Like NYCTA, for bus routes that are not interlined, the daily level data are thought to be reasonably reliable. Occasionally, however, weekend data are reported for routes that are not operated during weekends. There are two possible explanations for this. First, the driver may have entered the wrong route number. Second, since the time clock changes the date at midnight, some of the late night trips on Friday (early Saturday morning) may be classified as Saturday trips. The second problem was fixed by making the time clock date change occur at 3:30 a.m. instead of midnight.

WMATA experiences three common driver entry errors. First, the driver does not enter anything into the farebox. In this case, the daily revenue data can be retrieved at the garage level, but the ridership data are completely lost. Second, the driver does not register before some trips. In this case, some of the revenue and ridership information will be attributed to other trips. That is, the ridership and revenue information of some other trips will be inflated by the data from the missing trips. Finally, the driver logs on and off too frequently. Some drivers log on and off more than once per trip. In this case, the daily level revenue data are not affected, but the trip level data need to be handled carefully if they are to be retrieved correctly.

WMATA believes that driver entry errors occur for three major reasons. First, driver training in the use of electronic farebox is not uniform throughout the system. Although a clearly written manual is distributed to each driver, the amount of time spent on training and the teaching methods differ across instructors. Second, although some retraining is currently available for drivers, it is not working very well. These retraining sessions involve going through the whole data entry process with drivers and pointing out some of the most frequent driver entry errors. Drivers who have been identified as having problems with the electronic farebox are required to undergo retraining, but identification is not systematic. Finally, since some operators do not know the significance of entering the data correctly, they do not always do it carefully.

Regarding the problem operators, the Planning Department does try to identify them, and disciplinary actions may be taken for the worst cases. In general, the garage is notified of the problem operators, and is responsible for taking disciplinary action. Although there are many driver entry errors, WMATA feels that the operators are doing a reasonably good job given the number of buttons they have to press every day (operators have to press a button even for recording a full fare). However, even better performance is desired.

2.3.4 Seattle Metro Transit Authority (Seattle Metro)

Seattle Metro also experiences considerable driver compliance problems. Approximately 11% of total daily revenue is generally unclassified, and approximately

5% of trip records are usually associated with invalid signcodes. Drivers often neglect to reset the farebox at the beginning of each trip, and boarding counts can accumulate over multiple trips on multiple routes for interlined buses. However, no methods are currently employed to encourage driver compliance. Another problem is that each of the garages are restricted to a set number of valid signcodes that may be recognized by the data manager. Since this number is smaller than the total number of valid signcodes, significant problems arise when a vehicle from one garage covers a route that is assigned to another garage. In such a circumstance, all of the ridership and revenue data collected on-board may be allocated as unclassified once downloaded to the data manager.

2.3.5 Greater Cleveland Regional Transit Authority (Cleveland RTA)

Cleveland RTA also experiences driver non-compliance. One reason for this is the weak relationship between the Revenue and Operations Departments. Although drivers are occasionally reminded in general terms by Operations personnel to follow proper keypad entry procedures, this results in little more than "spurts of compliance." In addition, the agency also experiences a farebox problem in which the on-board hardware sometimes is unable to distinguish between different coins. In order to simplify the keypad entry procedures, fareboxes are programmed such that if a signcode is entered correctly for an outbound trip, the driver needs only to press '# *' to signify the beginning of the return trip.

2.3.6 New Orleans Regional Transit Authority (New Orleans RTA)

New Orleans RTA estimates the driver compliance rate to be about 90%. Despite some buses in the system being interlined and ten-cent transfers being available to passengers, the electronic farebox data are still good. In order to maintain high driver compliance, the agency takes a number of measures. First, driver compliance is monitored daily through electronic farebox printouts, and all drivers are evaluated every 3 to 4 months. Second, drivers who have made three major entry errors (such as failed to log on or incorrect entry of run number, driver number, or signcode) are subject to suspension. Third, training in the use of electronic farebox is provided to all new drivers as part of the 6-week comprehensive training program, with the average amount of time spent on the electronic farebox about 2 hours. Finally, retraining is provided for problem drivers. It is generally provided by an inspector either before the driver pulls out of the garage or during a trip, with the average amount of time spent on each retraining session about 20 minutes.

The agency also says that the electronic fareboxes consistently undercount both revenue and ridership. In general, the daily and monthly ridership information by route is low by about 5% while the revenue information at the same levels of detail is low by about 3%. The electronic farebox also undercounts annual ridership by route by about 6% to 7% and annual revenue by route by about 5%.

2.3.7 Phoenix Regional Public Transportation Authority (Phoenix RPTA)

Phoenix RPTA is quite successful with the electronic farebox despite the fact that on Valley Metro buses, operators must be especially conscientious about interacting with the fareboxes since special keypad entries must be made each time a bus crosses a city line (individual cities are assessed for costs of transit services according to vehicle miles and ridership within their boundaries). The agency says that 90% of the trips are correctly recorded. A correct record means that the operator ID, the route number, and the fareset are all valid as recorded by the farebox.

2.3.8 Orange County Transportation Authority (OCTA)

Although no quantitative figures are available, OCTA feels that most drivers are using the electronic fareboxes correctly, and driver compliance is not a problem. The one problem that is recurring is compliance on interlined routes with the drivers not entering the correct route number for each trip. Fortunately, OCTA does not do much interlining.

Farebox training for new drivers includes 2 to 3 hours classroom instruction on farebox operation. Additionally, each driver has about 40 hours of behind-the-wheel training that includes farebox usage. OCTA has no retraining for existing operators, however. The field supervision staff evaluates each driver's performance annually including use of the farebox, defensive driving, and customer service.

The OCTA Revenue Department does compare revenue figures. The difference between the GFI computer revenue total and the actual revenue counted averages 1% of the total with the actual revenue counted the higher of the two numbers.

2.4 Use for Planning and Route Evaluation

All surveyed agencies are using the electronic farebox for planning or performance evaluation to some extent; however, the kind of information used varies across agencies. In general, agencies which have satisfactory driver compliance (such as CTA and New Orleans RTA) tend to use both revenue and ridership information estimated from the electronic farebox whereas agencies which have less satisfactory driver compliance such as NYCTA and WMATA tend to use primarily revenue information estimated from the electronic farebox. None of the surveyed agencies are currently using the electronic farebox to estimate maximum passenger load information.

2.4.1 New York City Transit Authority (NYCTA)

Since 1990, NYCTA has been using daily electronic farebox (first GFI, later Cubic) revenue information by route for planning and evaluation purposes, but electronic farebox ridership information is not generally used yet due to low driver compliance, even though it is still collected at the route and day level. Instead, ride checks are used to gather ridership information. However, senior citizen ridership and electronic farecard usage are being obtained from the farebox. NYCTA states that there is potential for the revenue and ridership information to be collected at the route and direction level and the route and trip level. Metro Card usage, cash, tokens, and all types of fares are recorded.

NYCTA is also using the electronic farebox data for service evaluation. For example, the revenue information is used to evaluate NYCTA's competition with vans.

2.4.2 Chicago Transit Authority (CTA)

The electronic fareboxes have changed the way ridership information is collected at the CTA. Before electronic fareboxes, ridership information was estimated from revenue information. Now, the two types of information are collected separately with the electronic fareboxes also used for ridership reporting purposes.

CTA generally summarizes both ridership and revenue information at the route and hour level for planning purposes with weekends and weekdays handled separately. However, occasionally trip level information is also collected. There are only two cash fare types, full fare and reduced fare, in the system. There are hourly stamps on the data; that is, there is a separate record for all activities in each hour. The bus numbers, route numbers, farebox numbers, dates, faresets, driver numbers, and run numbers are also recorded. Before the use of the hourly time stamps, information was collected over five time periods of the day.

The CTA is also using the electronic farebox information to compute several evaluation measures. For example, quarterly reports are developed for all routes including farebox recovery ratios and average revenue per passenger. Routes in the top and bottom 20% in terms of performance are identified.

2.4.3 Washington Metropolitan Area Transit Authority (WMATA)

The GFI fareboxes were introduced in 1987, but were not used for planning purposes until 1992. The original purpose was to facilitate revenue capture and to reduce revenue loss thus revenue information was the primary focus, and electronic farebox ridership information is only now becoming important. In general, ride checks are still considered the primary tool for data collection while the electronic farebox is considered secondary. Both daily revenue and ridership information are collected by route and by garage, but only the revenue information is used directly at the route and the garage level, whereas ridership information is further summarized by jurisdiction. Generally, the weekday ridership data for each month are averaged by day of week, then the highest average of each month is picked out to compare with the highest weekday average of the same month for a previous period to show the ridership trends resulting from service changes, seasonality, or other effects. In addition, farebox data are adjusted or replaced with data from a comparable day if necessary. Currently, all types of fares are recorded on the GFI fareboxes.

2.4.4 Seattle Metro Transit Authority (Seattle Metro)

Seattle Metro's Planning Department has very little confidence in the electronic farebox data. As a result, it is not used to any significant extent for service planning. However, the Seattle Metro's Director of Planning has, on occasion, requested daily ridership data at the system-wide level. Usually, such data are used only to compare holiday ridership figures with those of regular weekdays in order to determine appropriate levels of holiday service.

2.4.5 Greater Cleveland Regional Transit Authority (Cleveland RTA)

The Strategic Planning Department produces quarterly Bus Route Performance Evaluations for most services. These evaluations group each of the routes according to primary operational characteristics (express, rapid transit feeder, etc.), and then rank them based on three separate measures of effectiveness. The raw ridership figures used for calculating these indices are provided solely by the electronic fareboxes.

2.4.6 New Orleans Regional Transit Authority (New Orleans RTA)

In New Orleans RTA, daily (and monthly) ridership and revenue information are estimated by route from the electronic fareboxes. In addition, all kinds of riders such as full fare, pass, student, senior citizen, and handicapped are recorded. The garage level revenue count serves as the control total for revenue estimation, and the ridership figure estimated from revenue serves as the control total for ridership estimation. While the full fare for the system is $1.00, a system level average fare of 68 cents is used for ridership estimation from revenue. The agency uses the electronic farebox data for both planning and annual service evaluation.

2.4.7 Phoenix Regional Public Transportation Authority (Phoenix RPTA)

Phoenix Transit produces quarterly Bus Route Performance Evaluations which rank each route according to standard measures of effectiveness, with the ridership figures coming from the electronic fareboxes. If a service change is to be implemented though, manual ride checks are performed for use in final evaluations.

2.4.8 Orange County Transportation Authority (OCTA)

OCTA collects daily revenue information by route and ridership information by route, day, and service hour. In addition, pass and coupon usage is recorded. Since only one button on the farebox is available for passes, the farebox does not distinguish between different types of passes whereas several buttons are available for coupons, so the farebox can distinguish one type of coupon from another. The agency uses the electronic farebox data and on-board ride checks for monthly service evaluation. It also uses ridership information for some Section 15 reporting purposes.

The Transportation Planning Department produces monthly ridership and productivity (passengers per vehicle service hour) reports for all fixed route bus service. It is split into eight modes: large local bus routes, small local bus routes, large express routes, small express routes, large rail feeder routes, small contract rail feeder routes, small contract local bus routes, and small contract express bus routes.

2.5 Agency Expectations

Many transit agencies have expectations of EF technology which are yet to be fulfilled. However, OCTA, New Orleans RTA, and Phoenix RPTA are very satisfied with their electronic fareboxes and currently do not have any major plans to improve the performance of the electronic fareboxes.

Since NYCTA is just starting to use the Cubic farebox information for planning purposes, it does not know how good the information will be, and it does not know exactly what to expect. However, the agency has two major objectives. First, by the middle of 1997, NYCTA fare collection is planned to be completely transformed to an electronic system. At that time, Metro Cards will be accepted on all buses and subway stations, and will eventually replace tokens. Second, NYCTA is planning to use the electronic farebox to collect information by route and time period (for 5 time periods), to handle transfers, and to provide universal transfers between buses and subways. Besides these two major objectives, a minor objective is that once the employee farecards with unlimited free rides are distributed, employee usage of the subway system will be tracked to prevent fraudulent employee pass use.

Although CTA is relatively successful with the electronic farebox, it has some goals which are not yet fulfilled. Specifically, CTA wants to accomplish three tasks in the near future. First, it is planning to produce driver exception reports to identify drivers who do not use the fareboxes correctly. Second, CTA is planning to use the planned Automatic Passenger Counting System (APC) to validate the electronic farebox data. Finally, CTA is planning to develop a new fare collection system which integrates the bus and rail systems. However, CTA is not thinking about integrating the GFI fareboxes with the destination signs at this point.

WMATA has two goals for the electronic fareboxes. The first one, which is very similar to what Phoenix RPTA is currently doing, is to have a reliable tool for collecting trip level information so that ride checks can be used more sparingly. The agency hopes to be able to use the electronic farebox data to observe revenue and ridership trends for each route. Then, if there is significant change in revenue or ridership for a route, ride checks will be performed to get more information. On the other hand, if there is no significant change in revenue or ridership for a route, ride checks will not be performed immediately. Therefore, the goal would be to postpone the ride checks for routes without significant change in revenue or ridership, and perform ride checks for routes that need immediate further study as signaled by significant change in revenue and ridership in the electronic farebox data. The second goal is to make information easily accessible at different levels of detail (monthly, quarterly, annually). In general, the GFI fareboxes are expected to fulfill varying data needs and provide flexibility in information access.

Although Seattle Metro and Cleveland RTA are not very satisfied with the performance of their electronic fareboxes, they do not have concrete plans yet for improvement. However, Cleveland RTA feels that the relationship between the Revenue and Operations Departments needs to be strengthened to achieve better driver compliance with the proper keypad entry procedures.

2.6 Validation of Electronic Farebox Data

All agencies except CTA and OCTA validate their electronic farebox data to some extent. In general, the agencies use garage revenue counts to validate the revenue data, and ride checks to validate the ridership data.

2.6.1 New York City Transit Authority (NYCTA)

In NYCTA, revenue counts by depot and month are used to validate revenue information from the electronic farebox. Since the revenue count is performed only 3 to 4 times a week in each depot, it is hard to gather daily revenue information by depot. The revenue count figure by depot and month is compared with the electronic farebox figure, with the result that the electronic farebox consistently undercounts revenue, generally by around 2%. Thus, the daily electronic farebox revenue figures by route need to be factored up to estimate the true revenue.

The ridership information gathered from the Cubic fareboxes is not used extensively yet, so there is no rigorous validation procedure for electronic farebox ridership data. However, ridership comparisons between ride check data and the electronic farebox data are just starting to be made. Daily ridership information by borough is estimated by the following equations:

Total ridership = Senior citizen ridership + Full fare ridership + Student ridership

Student revenue = Student ridership * Student fare

Senior Citizen revenue = Senior Citizen ridership * Senior Citizen fare

Full fare revenue = Total revenue - Student revenue - Senior Citizen revenue

Full fare ridership = Full fare revenue / Full fare

Senior citizen ridership is obtained from the electronic fareboxes, and student ridership is estimated from the number of student passes distributed. Since it is very difficult to estimate daily student ridership at the route level, all information mentioned above except senior citizen ridership is dealt with at the borough and day level. However, MetroCard will be introduced for all students in September 1996 for all buses; at that time, student ridership will no longer need to be estimated from the number of student passes distributed.

Although ridership information gathered by the electronic farebox is not used for planning purposes, ride check ridership data are compared with electronic farebox ridership data to get an idea of the quality of the electronic farebox data. The electronic farebox presumably undercounts ridership, but the magnitude of the undercount is not yet known.

2.6.2 Chicago Transit Authority (CTA)

In CTA, ride checks were done at the beginning explicitly to validate electronic farebox data, but they are now no longer done just for validation purposes. The match between the information collected by ride checkers and the information provided by the electronic fareboxes was very good. However, the number of ride checkers has not been reduced with the introduction of the electronic fareboxes. There are about 20 traffic checkers in CTA currently. Ride checks are still performed for Section 15 reporting purposes on items other than revenue and ridership as well as for peak load monitoring.

2.6.3 Washington Metropolitan Area Transit Authority (WMATA)

In WMATA, ride checks are used to validate electronic farebox ridership data. The electronic farebox trip level ridership information is compared with the ride check trip level ridership information. Currently, ride checkers collect ridership information for all routes, and the electronic fareboxes are not used to replace ride checks.

2.6.4 Seattle Metro Transit Authority (Seattle Metro)

Security inspectors have been deployed in the field to observe the number of passengers who overpay or underpay. With this information, Seattle Metro can better estimate a correct distribution of unclassified revenue. The electronic fareboxes are primarily viewed by Metro Operations as a tool to assist drivers in resolving fare disputes on-the-spot, and alerting security personnel to routes which have disproportionately high levels of unclassified revenue or unauthorized non-paying passengers.

2.6.5 Greater Cleveland Regional Transit Authority (Cleveland RTA)

Cleveland RTA recently deployed a number of APC-equipped coaches which produced passenger counts that are being compared with the electronic farebox readings. These APCs were provided by an independent contractor performing a comprehensive operational analysis -- since that project is now complete, the APCs have been removed from all coaches. However, electronic farebox boarding count validation is also accomplished through regular ridechecks on each route.

2.6.6 New Orleans Regional Transit Authority (New Orleans RTA)

In New Orleans RTA, ridechecks are performed for all routes and all modes to validate electronic farebox ridership data and to gather information for Section 15 reporting purposes. Although the match between the ridership information estimated from the electronic farebox and the ridership information gathered from ridechecks varies, it is consistently above 90%.

2.6.7 Phoenix Regional Public Transportation Authority (Phoenix RPTA)

In Phoenix RPTA, electronic farebox boarding count validation is accomplished through ride checks; however, it is performed relatively infrequently.

2.6.8 Orange County Transportation Authority (OCTA)

The agency does not validate the electronic farebox data at all because past experience, based on ride checks, shows the farebox data to be fairly accurate. In fact, the agency does not adjust the electronic farebox data to estimate revenue or ridership information; it simply uses the electronic farebox data directly. Ride checks are made to record running times, boarding counts, and alighting counts, but they are not used to validate electronic farebox data.

OCTA's current policy is to have each route checked once per year. Prior to October 1995, ride checks had not been done for 1.5 years due to budget cuts.

2.7 Conclusions

From the agency surveys, six major conclusions can be drawn. First, transit agencies which are reasonably successful with electronic fareboxes are those that are either very experienced with them or have a good training, retraining, and disciplinary program for operators. CTA and Phoenix RPTA, both having 10 years (or more) experience with electronic fareboxes, exemplify the first group while New Orleans RTA exemplifies the second. Second, most transit agencies use electronic farebox revenue data for planning or service evaluation purposes, but only those which are more successful with the electronic fareboxes (such as CTA, OCTA, Phoenix RPTA, and New Orleans RTA) also use electronic farebox ridership data for the same purposes. Third, most agencies (except OCTA) adjust the electronic farebox data to estimate the true revenue or ridership. Generally, the garage revenue count is used as the control total for estimating revenue information by route, and ride checks are used to validate electronic farebox ridership information. Fourth, none of the interviewed agencies are currently using electronic farebox data to estimate maximum passenger load. Fifth, some agencies such as WMATA and Phoenix RPTA are beginning to use electronic farebox ridership data to complement ride checks. Generally, these agencies would first look at electronic farebox ridership data to see if there has been any significant change in ridership, then if there is, ride checks would be performed to confirm the change before any action is taken. Finally, many agencies have goals that are not yet fulfilled. These goals can usually be classified as three types. The first type is to integrate the bus and the subway fare systems, the second type is to have better and faster access to different levels of detail of revenue and ridership information, and the third type is to improve driver compliance with the proper keypad entry procedures.

Chapter 3

Proposed Methods for Using Electronic Fareboxes in Data Collection

This chapter discusses the common types of farebox errors and presents the proposed methods, which are designed to overcome these errors, to estimate revenue and ridership information at different levels of detail. For the estimation of daily fare revenue by route, four methods are proposed. Method 1 allocates the unknown fare revenue to each route in proportion to their recorded fare revenue. Method 2 allocates the unknown fare revenue to each route based on missing trip records. Method 3 screens out bad trip records and then applies Method 2. Method 4 combines Methods 1 and 3. For the estimation of pass swipes of different kinds of passes by route, three methods are proposed (similar to the first three proposed methods for estimating daily fare revenue by route). The fourth method is not applied to daily pass swipes by route because Method 3 performs so well (see Chapter 4) that Method 4 is not needed. For the estimation of daily route ridership, a method which does not depend on a control total for ridership is proposed. This method depends on the definition of average fare and total pass swipes and uses Method 3 for estimating daily route fare revenue. For the estimation of route ridership by time period, a two-part method is proposed. The first part involves estimating the daily route ridership by applying the previous method. The second part allocates this ridership across the five periods of the day with the daily ridership serving as the control total.

Section 3.1 briefly discusses the different types of data that can be collected by electronic fareboxes. Section 3.2 discusses the common types of farebox errors. Section 3.3 presents the four methods for estimating daily route fare revenue. Section 3.4 presents the three methods for estimating daily pass swipes by route. Section 3.5 presents the method for estimating daily route ridership. Finally, section 3.6 presents the method for estimating route ridership by time period.

Many of the MBTA reports referred to in this chapter are shown in Appendix A of this thesis.

3.1 Data Collected from Electronic Fareboxes

Information can either be obtained directly from the fareboxes, or estimated from farebox data. Information can also be obtained at different levels of detail. Each of these topics is discussed below.

There are three major types of data that can be obtained directly from the electronic fareboxes; however, they have different levels of reliability. Furthermore, the different levels of detail also imply different levels of reliability.

3.1.1 Revenue Data

The hypothetically most reliable type of directly obtainable data is fare revenue. However, the information of interest at the route level is total revenue which includes both fare revenue and pass revenue. Since pass revenue cannot be obtained directly, it must be estimated from the number of pass swipes for each type of pass and the revenue per pass swipe for each type of pass. Fare revenue and the number of pass swipes for each type of pass can be obtained directly from the revenue and ridership summary printouts. However, first current revenue and unclassified revenue have to be defined. Current revenue includes both classified revenue, which is revenue that can be classified to specific fare categories, and unclassified revenue, which is revenue that cannot be classified to specific fare categories. The only drawback is that there are always undercounts in revenue, which will necessitate some form of farebox adjustment. The main cause for undercount is that some operators do not log on to the farebox properly at the beginning of each trip, which means the electronic fareboxes cannot associate the fare revenue and ridership for these trips with the appropriate routes. Instead, these fare revenue and ridership are classified as "others". In addition, if a farebox has more than 100 records between probes, only the data from the first 100 records will be recorded properly. The additional data are classified as "unknown". This scenario also applies to pass swipes and ridership data. Section 3.2 discusses unknown and other revenues in slightly more details. Therefore, information must always be estimated rather than obtained directly from the farebox report summaries.

Information can be obtained at four different levels of detail. Fare revenue at the garage and day level is known with certainty and can be thought of as a control total. It can be obtained from the revenue and ridership summary by route printouts. The only exception is for buses which are not vaulted on some days; this is especially common for small garages where resources are limited. This control total includes unknown and other revenues. Although information at the garage and day level is not very useful by itself, it is hypothetically the most reliable information, and it can be used to estimate information at lower and more useful levels of detail such as by route.

The next hypothetically most reliable data is by route and day. It can also be obtained from the revenue and ridership summary by route printouts, but once again there are also unknown and other revenues. The next hypothetically most reliable data is by route and time period. Finally, the least reliable data is by route and trip. It can also be obtained from the revenue and ridership summary printouts, but some trips are likely to be missing. Generally the reliability is expected to decrease as the level of detail is increased because some errors which will affect the more detailed data will not affect the more aggregate data, at least to the same extent.

Pass swipes are generally considered to be less reliable than fare revenue because some passengers simply flash their passes at the drivers or do not swipe their passes properly through the pass readers. Either way, these passengers and their pass usages are not recorded by the electronic fareboxes. Therefore, the number of pass swipes is likely to underestimate the number of pass riders. Unlike the improper registration of fare revenue, where the collected fare revenue is still physically in the farebox even though it may be recorded as "others" or "unknown", these pass swipes are lost forever. Hence fare revenue is generally considered to be more reliable than pass swipes.

3.1.2 Ridership Data

The next most reliable kind of directly obtainable information is the number of passenger boardings. It can also be obtained from the revenue and ridership summary printouts. However, the quality of information varies with the data entry behavior of each driver, and there are also undercounts, which require the use of a farebox adjustment factor. Undercounts are generally due to operators not registering special fares properly. For example, if two student riders deposit two half fares into the farebox, the farebox considers these two half fares as a full fare unless the operator explicitly registers them as two half fares by pressing the correct buttons on the farebox. This is why ridership data are generally considered to be less reliable than revenue data. In addition, unlike fare revenue data, there are no control totals available for ridership, which makes the estimation of ridership more difficult than that of fare revenue. Alternatively, passenger information can also be estimated from revenue data through the use of an average fare.

As with revenue information, the reliability of the ridership data is expected to decrease as the level of detail is increased; i.e. the garage level will be the most reliable and the trip level will be the least reliable.

3.1.3 Other Usage Data

Some information such as passenger miles and maximum passenger load cannot be obtained directly from the electronic fareboxes, but they may be estimated based partially on farebox data. However, traffic checks, specifically ride checks, are required to develop relationships between particular kinds of directly obtainable information and particular kinds of estimated information. After that, conversion factors for these kinds of information can be developed. Passenger mile information and maximum passenger load information are expected to be less reliable than the passenger (from revenue data) information. Maximum passenger load information is not studied in detail in this research because the first test is to estimate the basic information-revenue and passengers; if the results are favorable, more complex and problematic data types like maximum passenger load and passenger miles can be studied later.

3.1.4 The Emphasis of this Research

Given that our objective is to determine at what level reliable information can be obtained, this study starts with what is expected to be the most reliable kind of useful information of different types. If the reliability proves to be acceptable, a finer level of detail for the same kind of information can be investigated. If information at a particular level of detail does not prove to be reliable, efforts should be made to identify and correct the problems before proceeding to finer levels of detail.

In this thesis, the emphasis will be on the following four items: daily fare revenue by route, daily pass swipes by route (for estimating daily pass revenue by route), daily ridership information by route, and ridership by route and time period (for use directly as well as for potentially estimating maximum passenger load by route and time period).

These items are emphasized because they are most useful for planning purposes and also the most promising basic data types.

Table 3.1 shows a matrix relating the level of detail to the type of information.

Table 3.1: Matrix Relating the Level of Detail to the Type of Information

Type of Info / Level of Detail	Fare Revenue	Pass Usage	Ridership
Daily Route Level	***	***	***
By Route and Time Period			***
By Route and Trip			

*** Items that will be emphasized.

For each item to be estimated in this study, it is generally emphasized at the least level(s) of detail. As mentioned before, the reason is that our objective is to determine at what level reliable information can be obtained, so this study starts from the least level(s) of detail and proceeds to finer levels of detail if the results are favorable.

3.2 Common Farebox Errors

This section discusses the common types of farebox errors. There are three major types of errors which may cause electronic farebox data to be inaccurate. The first type is non-operator-related error including errors that are beyond the control of the operator such as farebox or pass reader defects. The second type is operator-related error including errors that are caused by the operator. Some examples of operator-related errors are neglecting to reset the farebox after each one-way trip; overusing the dump button; entering incorrect signcodes, run numbers, or badge numbers; logging on more than once for each trip; not pressing the appropriate button on the farebox when a passenger boards. Each of these errors can cause some of the electronic farebox data to be inaccurate. The third type of errors are those that are external to the electronic fareboxes, they are usually related to operational practices which can affect fare collection. Some examples are back-door boarding, flashing instead of swiping of passes, failure to vault the fareboxes at the end of a day. In these situations, the revenue and ridership cannot be captured by the electronic farebox system, and are therefore very difficult to account for. Since these are operational problems not directly related to the farebox itself, the methods presented in this thesis do not attempt to solve these problems. However, these operational problems may still affect the validity of the information of interest, and they must be resolved if the information is to be useful for planning purposes. Small routes generally cannot tolerate errors as well as larger ones because they generally have fewer trip records, and if errors do occur, there may not be enough good records remaining to estimate the desired information reliably.

Farebox defects can cause either revenue or ridership (or both) to be significantly undercounted. In the worst case, none of the revenue or riders are counted. In this case, the electronic farebox is essentially a mechanical non-registering farebox. However, the total amount of fare revenue collected can still be obtained from the garage level fare revenue count since the fare revenue is still collected, though not registered, by the electronic farebox. Similarly, pass reader defects can cause the number of pass swipes to be significantly undercounted. However, the unregistered pass swipes cannot be recovered because no garage level control total exists for pass swipes.

Neglecting to reset the farebox after each one-way trip can cause the revenue and ridership data of the first trip of the series to be inflated because the data of the first trip include data from all subsequent trips until the operator resets the farebox. Therefore, the trip level data can be seriously affected. Furthermore, if the bus is interlined (operated on several routes), some of the recorded revenue and ridership of a route may actually belong to another route, and therefore the route level data can also be in error.

Overusing the dump button can result in a high proportion of revenue being unclassified and ridership being undercounted. The intended use of the dump button is to dump any overpayment into the farebox as unclassified revenue. Unfortunately, some operators may use the dump button even for legitimate special fares such as senior citizen and student fares. In this case, these legitimate special fares are considered unclassified revenue and the riders are not counted. Therefore, a high proportion of revenue will be unclassified and the ridership undercounted.

Entering incorrect signcodes can cause the associated revenue and ridership to be counted as "others" because the farebox cannot allocate the revenue and ridership to a route. Unknown fare revenue is the revenue that cannot be allocated to any route or signcode because of farebox data overflow. Usually a farebox can hold 100 records, if the number of records exceeds 100 before the farebox is probed, anything beyond the 100th record will be classified as unknown fare revenue. Similarly, any ridership beyond the

100th trip will be classified as unknown ridership. Fare revenue that cannot be allocated to any route or signcode because the operator does not register with the farebox properly is classified as "others". However, for simplicity, this thesis will define both "others" and unknown fare revenue simply as unknown fare revenue.

Entering incorrect run numbers or badge numbers affects the quality of the trip level data because it is difficult to identify a trip if the run number or badge number is incorrect.

Logging on more than once for each trip will result in extra (partial) trip records. This can significantly affect trip level data because both revenue and ridership are undercounted as some of the revenue and riders are considered to be associated with the "extra" trips. However, the route level data should generally not be affected.

If an operator does not press the appropriate button when a special-fare passenger boards, the ridership data can be affected. For example, as mentioned before, if two half-fare passengers board, and the operator does not press the appropriate buttons, the electronic farebox will assume that one full-fare passenger has boarded. However, the revenue data should generally not be affected by this error.

The problems of overusing the dump button and entering incorrect signcodes generally should not seriously affect the fare revenue control total because the miscounted fare revenue should either be considered as "others" or be counted as that of another route(s), which are still part of the control total. Therefore, these problems will be dealt with in Methods 1 and 2, and the allocation part of Methods 3 and 4. The problem of not pressing the appropriate button when a passenger boards affects ridership data but not revenue data, so it is dealt with by estimating ridership information based on revenue information through an average fare instead of electronic farebox ridership data. The three remaining farebox problems of defective fareboxes, neglecting to reset the farebox after each one-way trip, and logging on more than once for each trip will be dealt with directly in the screening part of Methods 3 and 4. Although, as mentioned in the previous paragraph, the problem of logging on more than once for each trip should not generally affect route level data, since Methods 2 through 4 allocate revenue and ridership based on the number of trip level records, and this problem affects the number of "missing" trip records, this problem will be dealt with in this thesis.

3.3 Daily Route Fare Revenue

Four different methods are proposed for estimating daily route fare revenue. All four methods focus on the problem of how to best allocate the unknown fare revenue across routes. In addition, Methods 3 and 4 attempt to screen out bad trips before performing the allocation of unknown fare revenue. The four methods are:

- Method 1: Proportional Allocation of Unknown Fare Revenue
- Method 2: Allocation based on Missing Trip Records
- Method 3: Screen Bad Trip Records then Apply Method 2
- Method 4: Combination of Methods 1 and 3

3.3.1 Method 1: Proportional Allocation of Unknown Fare Revenue

Method 1 allocates the unknown fare revenue to each route in proportion to the recorded fare revenue of that route. Therefore, routes with large recorded fare revenue are allocated more of the unknown revenue than routes with smaller recorded fare revenue. The method is based on the hypothesis that the effects of driver entry errors are uniform across all routes in the garage. Only the signcode and the route revenue fields of the Revenue and Ridership Summary by Route are needed for this method (see Appendix A).

The method consists of the following three simple steps:

1. Obtain the total fare revenue (including unknown fare revenue) for the garage for the day of interest from the Revenue and Ridership Summary by Route. This is the garage/

day control total to which all final route revenues must sum. The control total C is simply the sum of R_i, the recorded fare revenue for route i in the garage for the day of interest, and U, the unknown fare revenue for the garage for that day.

2. Determine the factor F by dividing the control total, C, by the sum of R_i.

3. Finally, the recorded fare revenue of each route in the garage for the day of interest is factored up by F to obtain the estimated fare revenue for that route for that day.

3.3.2 Method 2: Allocation Based on Missing Trip Records

Method 2 allocates the unknown fare revenue to each route considering both the fare revenue recorded for each route, and also the discrepancy between the number of recorded trips and the number of scheduled trips on each route. Therefore, routes with large recorded fare revenue and many missing trips are allocated more of the unknown fare revenue than those with smaller recorded fare revenue and fewer missing trips.

This method is based on two hypotheses:

1. The missing trips for each route generate an amount of fare revenue per trip equal to the average of all the recorded trips for that route.

2. All missing fare revenue becomes unknown fare revenue. That is, it does not consider explicitly the possibility that some missing fare revenue gets counted as recorded fare revenue of another trip on the same route or on another route.

This method uses the signcode, number of runs, and route revenue fields of the Revenue and Ridership Summary by Route and the number of scheduled trips for each route.

This method consists of three steps:

1. Estimate the missing fare revenue of each route in the garage for the day of interest by multiplying the mean revenue per trip for route i (R_i/RT_i) by the number of missing trip records on that route-simply the difference between the number of scheduled trips ST_i and the number of recorded trips RT_i.[1]

2. Allocate the unknown fare revenue for the garage to routes in proportion to the

---

1. If the number of recorded trips exceeds the number of scheduled trips, the number of missing trips is set to zero.

missing revenue at the route level, MR_i, estimated in step 1 above.

3. The estimated fare revenue for each route is calculated by adding the recorded fare revenue for that route and its allocated fare revenue AR_i.

3.3.3 Trip-Level Screening

This section describes the trip-level screening that is used by both Methods 3 and 4 for estimating daily route revenue.

Table 3.2 shows the median inter-record time of each route in Bartlett Garage calculated from both 1994 data and 1996 data, that will be used for the first and second rounds of screening described below. Since the 10 days used for the pass swipe analysis are different from that used for the fare revenue analysis in 1996, different median inter-record times are generated. However, for simplicity, the combined medians are shown. Method 3 was tested with the combined medians, the results were very similar to those generated from using different medians for the fare revenue and pass swipe analyses. All results presented in this thesis are based on different medians for the fare revenue and pass swipe analyses. The range of book time (the trip time estimated by the MBTA) for each route is also shown. This range of book time is generally smaller than the median inter-record time because the median inter-record time also includes the recovery time. The last two columns show the percentage of trip records that have inter-record time less than 0.5 times the median and greater than 1.5 times the median, respectively. These percentages are based on combined 1994 and 1996 data with their own corresponding medians.

With the exception of routes 26 and 34E, the median trip times of all routes in 1994 do not differ significantly from those of 1996. The huge discrepancy in route 26 can be explained by the small percentage of trip records out of the number of scheduled trips and the large number of bad records in both 1994 and 1996. It is not unreasonable for the route

Table 3.2: Median Inter-Record Times

Route Number	Median Inter-record Time (1994)	Median Inter-record Time (1996)	Book Trip Time Range	% of records < 0.5 * median	% of records > 1.5 * median
14	30	30	25-28	4	23
24	23	23	11-22	7	29
26	73	31	19-26	14	63
27	16	15	11-12	2	16
28	41	39	24-41	11	31
29	35	32	24-36	8	36
30	24	26	9-14	7	32
31	20	20	11-18	7	23
32	24	23	12-21	6	24
33	28	28	22-26	18	27
34	26	26	13-23	6	31
34E	60	34	47-62	15	37
35	35	33	21-31	7	26
36	27	29	15-20	7	31
37	29	30	15-22	5	28
38	21	22	12-19	9	25
39	37	37	21-34	9	27
40	30	31	16-23	4	25
41	17	19	10-17	7	21
42	28	28	17-22	12	24
46	16	15	10-13	1	13
48	30	28	24	8	49
50	25	26	11-14	4	25
51	37	36	26-32	3	23
52	47	42	39-41	15	29

to have significantly inconsistent median trip times between 1994 and 1996. The cause of the discrepancy in route 34E is not so clear.

3.3.3.1 Median Inter-Record Times

The screening uses median inter-record time extensively. This is the median amount of time between two consecutive records of a bus on a route. The inter-record time has two components: trip time, the amount of time to complete a one-way trip, and recovery time, the amount of time between the completion of a one-way trip and the beginning of the return trip. Inter-record times are useful for determining if a trip record is legitimate. For example, if a trip record is chronologically very close (less than 0.5 time the median inter-record time for the route) to the following record, it is likely that a trip is registered multiple times by the operator. On the other hand, if a trip record is chronologically very far (more than 1.5 times the median inter-record time for the route) from the following record, it is likely that this record includes information from one (or more) following trips.

The median inter-record time instead of the average inter-record time is used to avoid bias due to statistical outliers. Although the bounds 0.5 and 1.5 are somewhat arbitrary, they are chosen to reduce the chance of eliminating a good trip record or including a bad trip record. For example, if 0.5 is increased to 0.9, there is a reasonable chance that some legitimate records will be screened out because of variation in trip times. On the other hand, if 0.5 is reduced to 0.2, it is likely that data from many multiple registered trips will be (erroneously) included. Similarly, if 1.5 is increased to 1.8, it is likely that some trip records with multiple trip data (inflated records) will not be screened out. On the other hand, if 1.5 is reduced to 1.1, there is a reasonable chance that legitimate trips will be screened out when the following trip departs slightly late. Table 3.3 shows the time histogram for route 28 based on 1994 data to illustrate the spread of inter-record times.

Table 3.3: Inter-record Time Histogram for Route 28 (1994 Data)

```
Histogram for Route 28

Median = 41.00
Trip Time < 10 Minutes                        113
Trip Time 10 Minutes to 19 Minutes            21
Trip Time 20 Minutes to 29 Minutes            67
Trip Time 30 Minutes to 39 Minutes            229
Trip Time 40 Minutes to 49 Minutes            234
Trip Time 50 Minutes to 59 Minutes            94
Trip Time 60 Minutes to 69 Minutes            44
Trip Time 70 Minutes to 79 Minutes            21
Trip Time 80 Minutes to 89 Minutes            25
Trip Time 90 Minutes to 99 Minutes            15
Trip Time 100 Minutes to 109 Minutes          10
Trip Time 110 Minutes to 119 Minutes          12
Trip Time >= 120 Minutes                      63
```

In this case about 47% of the trip records would have been treated as legitimate with the remaining 53% identified and treated as described below.

Four rounds of elimination screening have been developed in an attempt to eliminate the most common problems with the trip records:

3.3.3.2 Multiple records for single trip

The first round of elimination attempts to identify and deal with the problem of multiple registrations for a single trip. Multiple registrations mean that a driver registers more than once for a single trip, which creates multiple records for that trip with each containing only partial information. In order to identify trips with multiple records, the median inter-record time for each route is used. Since it is unlikely that a trip will be completed in less than half the median trip time for that route, this screening is based on the premise that if two trip records are separated by less than half the median trip time for that route, it is likely that the trip has multiple records. Therefore, when consecutive trip records (with the same bus number) occur in less than half the median inter-record time for that route, the revenues of these two trips are accumulated and recorded in a single trip record, and the earlier time is considered to be the time of the record because this is the time when the trip actually begins.

Table 3.4 contains sample data (ordered by signcode) which shows the problem of multiple registrations. From the figure, we can see that three records were created on bus 205 for signcode 280 at 12:48 and 12:49 on 3rd January 1996. The listed record times are when the trip record began rather than when they were completed. Since the bus numbers are identical in all three records (with recorded revenues $0.15, $0.00, and $16.33), and the three records are so close together (within two minutes while the median inter-record time for route 28 is 41 minutes and the range of book time is 24-41 minutes), it is highly likely that the three records reflect the same trip (probably the operator logged on three times during that trip to get the correct run number 1132). Therefore, in this round of screening the recorded revenues are accumulated ($16.48), and the three records are treated as a single record with the earliest time (12:48) considered to be the time of the new record. The two records created at 13:07 on bus 260 is another example of multiple registrations during a single trip. In this case, a new record is created with revenue $11.29 and the time 13:07 is considered to be the time the record was created.

Table 3.4: Sample Data Showing Multiple Registrations

Signcode	Run #	Bus #	Day	Month	Time	Revenue	Unclassified
280	132	205	03	JAN	12:48	0.15	0.00
280	11	205	03	JAN	12:49	0.00	0.00
280	1132	205	03	JAN	12:49	16.33	4.78
280	1116	207	03	JAN	13:01	40.59	7.79
280	1075	260	03	JAN	13:07	0.00	0.00
280	1065	260	03	JAN	13:07	11.29	0.51
280	1057	202	03	JAN	13:23	11.93	0.16

3.3.3.3 A single record covering multiple trips

The second round of screening attempts to identify instances of unreasonably high fare revenue. If an operator forgets to reset the farebox at the start of each trip, the fare revenue of all subsequent trips will be recorded as part of the first trip. This way, there is no record for all subsequent trips, but a very high (inflated) fare revenue recorded for the last trip registered correctly. This is highly undesirable because some of the trip record fare revenue actually belongs to another trip, and worse yet, even another route in the case of interlining. In this case, the number of recorded trips will be lower than the number of scheduled trips since some trips are not recorded, but the total recorded fare revenue may be no lower than the true fare revenue because all collected fare revenue will be associated with a single trip, which is still counted as the fare revenue for the route if interlining is not practiced.

This is problematic for Method 2 because an unreasonably large portion of the unknown fare revenue will be allocated to this route due to a high estimated "missing fare revenue" as a result of a large number of "missing trips". Of course, if interlining is practiced, the recorded fare revenue of the first trip contains fare revenue from other routes, which makes the problem even worse.

Median inter-record time is again used to determine if a trip is likely to contain inflated fare revenue. Since it is highly unlikely that a trip will take more than 1.5 times the median inter-record time for that route to complete, this round of screening hypothesizes that if the time between successive trip records is greater than 1.5 times the median inter-record time for that route, then the operator has not reset the farebox before the new trip, and hence the recorded fare revenue for that trip is inflated. Therefore, when consecutive trips (with the same bus number) are separated by more than 1.5 times the median inter-record time for that route, the first trip of the pair is excluded since its recorded fare revenue is likely to be inflated. Furthermore, since it is hard to determine the inter-record time of the last trip of each bus, it is also automatically excluded.

Table 3.5 contains sample data (ordered by bus number) which shows the problem of inflated fare revenue. From the figure, we can see that all trips made by bus 238 on route 32 on the 3rd of January in 1996, except the first trip, were completed in less than 1.5 times the median inter-record time of the route (35 minutes). The first record was created at 4:11, and the second was created at 10:22, which was more than 6 hours after the first one was created. This suggests that the operator probably did not reset the farebox after the first trip and that the fare revenue for the first trip record really reflects revenue for multiple trips. This suggestion is further supported by the fare revenue of the first trip record ($87.03), which is significantly higher than those of the other trip records. Therefore, the first trip is excluded as a result of this screening.

Table 3.5: Sample Data Showing Inflated Fare Revenue

Signcode	Run#	Bus#	Day	Month	Time	Revenue	Unclassified
320	1004	238	3	JAN	04:11	87.03	8.88
320	1016	238	3	JAN	10:22	3.55	0.00
321	1016	238	3	JAN	10:39	7.77	0.95
320	1016	238	3	JAN	11:07	2.50	0.10
321	1016	238	3	JAN	11:23	5.20	0.20
320	1016	238	3	JAN	11:53	1.70	0.00
321	1016	238	3	JAN	12:13	12.30	0.15
320	1076	238	3	JAN	12:46	8.45	1.25
321	1076	238	3	JAN	13:06	11.90	0.95
320	1076	238	3	JAN	13:35	9.52	0.52

3.3.3.4 Defective Fareboxes

The third and fourth rounds screen out trips made by buses that have defective electronic fareboxes. When a farebox records zero revenue for the whole day, it is considered defective for that day. By ordering records by bus number such buses can easily be identified by visual inspection. One possible cause of this situation is that a farebox is badly jammed, and the bypass knob is used, which disables all the registering functions of the electronic farebox.

Although this eliminates all buses with fareboxes with zero revenue for the whole day, other farebox defects may result in zero revenue for most, but not all, of the trips. This round assumes that a bus with five or more consecutive zero-revenue trips after all other screenings has a defective farebox, and hence all trips made by this bus are excluded.

Table 3.6 shows sample data (ordered by bus number) indicating the role of this fourth round of screening. In the figure, records with an asterisk in front of them are records that are eliminated in one of the first three rounds of screening. We can see that after these first three rounds, there are five consecutive zero-revenue records, made at the following times: 7:17, 7:40, 7:57, 8:18, and 9:25, indicated by an "X". This suggests that the electronic farebox on bus 239 was likely defective on the 5th of January, and so all records created on bus 239 on this day are eliminated.

Table 3.6: Sample Data Showing a Defective Farebox

	Signcode	Run #	Bus #	Day	Month	Time	Revenue	Unclassified
*	311	1088	239	5	JAN	06:37	0.00	0.00
X	310	1088	239	5	JAN	07:17	0.00	0.00
X	311	1088	239	5	JAN	07:40	0.00	0.00
X	310	1088	239	5	JAN	07:57	0.00	0.00
X	311	1088	239	5	JAN	08:18	0.00	0.00
*	310	1088	239	5	JAN	08:36	0.00	0.00
*	391	7	239	5	JAN	09:25	0.00	0.00
X	391	1088	239	5	JAN	09:25	0.00	0.00
*	390	1088	239	5	JAN	10:06	0.00	0.00
	280	9031	239	5	JAN	15:50	6.39	0.15
*	280	9031	239	5	JAN	16:35	0.00	0.00
	281	9031	239	5	JAN	16:35	5.30	1.45
*	280	9031	239	5	JAN	17:19	18.32	0.00
	281	9031	239	5	JAN	17:19	18.32	0.00
*	280	9031	239	5	JAN	18:19	0.00	0.00
*	281	9031	239	5	JAN	18:19	4.70	0.20
*	281	9031	239	5	JAN	19:43	0.00	0.00

3.3.4 Method 3: Screen Bad Trip Records then Apply Method 2

Method 3 introduces the concept of attempting to screen out bad trip records and then applying Method 2 to the reduced data set; however, both the number of trips and the split between classified and unknown data will be quite different as a result of the trip record level screening. The revenue (if any) associated with the screened out trips is added to the unknown revenue count for later allocation. The screening attempts to eliminate all bad trip records and thus recognizes the possibility that operator behavior will result in some data being wrong. However, this method still keeps the remaining hypotheses for Method 2. Furthermore, this method is conservative in the screening because it is better to eliminate records which are actually good than include bad records. Since this method is based on the trip level records in the Trip-Level Revenue and Ridership Summary (for the record screening) as well as the Revenue and Ridership Summary by Route, it is considerably more complicated and time-consuming than Methods 1 and 2.

Therefore, Method 3 consists of the following two steps:

1. Screen out bad records following the screening procedure described above.
2. Apply Method 2 to the data that are not screened out by Step 1 above.

Method 3 is theoretically better than Methods 1 and 2 because it attempts to screen out bad data, which is the key problem in this study. However, it also has two disadvantages. First, it is considerably more complex than Methods 1 and 2, which means a significantly longer processing time is required. Second, some small routes may end up with no data since all data are screened out. Method 4 below attempts to solve the second problem.

3.3.5 Method 4: Combination of Methods 1 and 3

Method 4 is a hybrid of Methods 1 and 3. Since Method 3 makes a great effort to identify and eliminate bad data, it suffers one potentially serious problem: complete elimination of all data for a route. This is especially likely with very small routes since they do not have a lot of data to start with, and some of them may have only one operator for an entire time period. If this operator does not use the proper farebox entry procedure, it is likely that all records this operator produces, and so all records for this small route, may be eliminated by Method 3. This problem is not so serious for the large routes because they have many records to start with and many operators on the route each day. Therefore one operator with poor compliance will not ruin the route data base for the whole day. In order to solve this problem, Method 4 applies Method 1 instead of Method 3 to routes that have very little or no data remaining after application of Method 3. Routes with sufficient data after the screenings are analyzed by Method 3 since it is more theoretically grounded than Method 1.

This method incorporates the hypothesis of Method 1 for the routes that are to be analyzed by Method 1, and the hypotheses of Method 3 for the remainder. It consists of five steps:

1. Apply the screening procedures of Method 3 to all routes.
2. Look at the number of trip records left (after the screening procedures) for each route.
3. Apply Method 1 to all routes which have less than 15 trip records left after the screening procedure. Basically this allocates some of the control total (and hence unknown fare revenue) to the routes which have less than 15 trip records left after the screening procedure.
4. Subtract the allocated fare revenue of the above routes from the garage control total for all ten days. This reduces the unknown fare revenue by the amount allocated to the routes which are analyzed by Method 1 in Step 3.
5. Apply Method 3 to the remaining routes based on the reduced control totals (and hence reduced unknown fare revenue).

3.4 Daily Pass Swipes by Route

As discussed previously, to estimate pass revenue at the route level it is necessary to estimate the number of pass swipes at the route level so that total pass sales revenue can be allocated to the route level. There are two differences between the estimation of fare revenue and the estimation of pass revenue. First, although garage level fare revenue counts are not used in the fare revenue analysis in the previous section because many fareboxes are not vaulted daily (the sum of daily fare revenue for all routes and the unknown daily fare revenue recorded by electronic fareboxes is used instead), they are nevertheless available to serve as control totals if garage level fare revenue counts are performed more regularly and become more reliable. However, no such control totals are available for pass revenue. Second, since passes are issued systemwide, the allocation of pass revenue will have to be across all routes in the system in order to get a true "revenue equivalence" for a pass swipe. However, only pass swipes (not pass revenue) will be analyzed in this thesis; the issue of estimating a pass swipe revenue equivalence is a separate question not directly related to electronic fareboxes.

Three methods are proposed for estimating daily pass swipes by route which closely parallel the first three methods proposed for estimating fare revenue as described in the previous section. Method 4 is dropped because Method 3 performs so well that Method 4 is not needed. Specifically, after the application of the screening procedure to the spring 1996 data, none of the routes has less than a total of 15 trip records remaining, which means that none of the routes needs to be analyzed by Method 1. In this case, Method 4 is identical to Method 3, and therefore is not needed.

In this thesis, two types of passes are analyzed: local bus and combo, as described below:

1. The local bus pass entitles the holder to unlimited travel on all local bus routes and the surface segments of the Green Line (except for the Riverside Line), and the holder may travel on zoned and express buses with the difference between the zoned or express bus fare and the local bus fare made up in cash.
2. The combo pass entitles the holder to unlimited travel on all local and zoned buses, subways (except Quincy Adams and Braintree stations of the Red Line), $1.50 express buses, and the local segments of the commuter rail lines. The holder may also travel on express buses with fares higher than $1.50 with the difference made up in cash.

Although there are several other types of passes (including the commuter rail passes), since they are not used frequently on bus routes, especially the mostly local and zoned ($1.00) bus routes that are operated out of Bartlett Garage (the garage on which the testing in chapter 4 is based), they are not analyzed in this thesis.

Both types of passes are analyzed using the same methods, but they are analyzed separately. All methods deal with the allocation of unknown pass swipes to routes and the errors affecting the records. Like the unknown fare revenue, unknown pass swipes (including "others" pass swipes) are pass swipes that cannot be allocated to a particular signcode either because the operator does not use the farebox properly, or because the farebox has more than 100 records.

3.4.1 Method 1: Allocate the Unknown Pass Swipes Proportionally

Method 1 allocates the unknown pass swipes of each pass type to each route in direct proportion to the recorded pass swipes of each pass type of each route. Routes with large number of recorded pass swipes are allocated more unknown pass swipes than those with fewer recorded pass swipes. This method hypothesizes that the effect of driver entry errors is uniform across all routes in the garage. This method is similar to Method 1 described in Section 3.3.1 following the same three steps with pass swipes replacing fare revenue.

3.4.2 Method 2: Allocate the unknown pass swipes based on missing trips

Method 2 allocates the unknown pass swipes for each pass type to each route based on the observed number of pass swipes per trip and the number of missing trips on the route. Therefore, routes with large numbers of recorded pass swipes and many unrecorded trips will be allocated more of the unknown pass swipes than those with smaller recorded pass swipes and fewer unrecorded trips. The method and the underlying hypotheses are identical to the Method 2 for fare revenue as described in Section 3.3.2.

3.4.3 Method 3: Screen for Bad Trip Records then Apply Method 2

Similar to Method 3 for estimating daily route fare revenue, Method 3 for estimating daily pass swipes by route also consists of a trip screening component and an allocation part. The pass swipes (if any) associated with the screened out trips are added to the unknown pass swipe count for later allocation.

As described in Section 3.3.3, there are four rounds of screening trip records:

The first and second rounds of screening eliminate multiple registrations and multiple trips included in a single trip record. This screening is identical to that described in Section 3.3.3 except that pass swipes instead of fare revenue are accumulated in the second round of screening.

The third and fourth rounds screen trips made by buses that have defective electronic farebox. When a farebox records zero pass swipes for local bus and combo passes for an entire day, it is assumed to be defective on that day. Table 3.7 contains sample data (ordered by bus number) which show a bus with such a defective farebox. We see that the bus has zero recorded pass swipes (both local bus and combo passes) for the whole day. In addition, we can see that the bus was operated by more than one operator on the 4th of February in 1994 because the records of this bus include three different run numbers. This means that it is highly unlikely that driver entry errors contributed to zero recorded pass swipes for the bus on that day. One possible explanation is that the pass reader on the farebox on bus 8673 was not functioning on that day, and so the farebox failed to record any pass swipes. Therefore, the farebox on this bus is considered defective on the 4th of February in 1994 with respect to pass data. The farebox of this bus is not automatically considered defective for fare revenue analysis. The reason is that the farebox may be registering cash fares correctly even though the pass reader is defective. Similarly, if a farebox is considered defective in the fare revenue analysis, it is not automatically considered defective in the pass analysis.

Table 3.7: Sample Data Showing a Defective Farebox

Signcode	Run#	Bus#	Day	Month	Time	TTP3	TTP5
320	9032	8673	4	FEB	00:00	0	0
280	1088	8673	4	FEB	06:34	0	0
280	1088	8673	4	FEB	08:11	0	0
280	1088	8673	4	FEB	08:11	0	0
391	9049	8673	4	FEB	15:25	0	0

The final screening assumes that a bus with five or more consecutive zero-pass-swipe trips after all other screening has a defective farebox, and hence all trips made by this bus will be excluded. Table 3.8 contains sample data (ordered by bus number) which shows the effect of this screening. In the figure, records with an asterisk have been eliminated in the first three rounds of screening. We can see that this leaves eight consecutive zero-pass-swipe records, made at the following times: 8:45, 9:18, 11:07, 11:31, 11:43, 11:57, 12:18, and 12:46, indicated by an "X". This suggests that the pass reader on bus 222 was likely defective on 2nd April 1996, and so all records created on bus 222 on this day are eliminated in this round of screening.

3.5 Daily Route Ridership

Although the methods for fare revenue and pass swipe analyses could be extended directly to ridership, this approach is not followed here because there are no obvious control totals for ridership at the garage level. Since the unknown ridership on the Revenue and Ridership Summary by Route, unlike unknown fare revenue, is not very accurate, we can no longer obtain the control totals for ridership at the garage level simply

Table 3.8: Sample Data Showing Defective Farebox

	Signcode	Run#	Bus#	Day	Month	Time	TTP3	TTP5
*	391	1012	222	2	APR	04:28	2	0
	390	1012	222	2	APR	05:30	0	0
	390	1012	222	2	APR	06:23	1	0
	391	1012	222	2	APR	06:52	1	0
	390	1012	222	2	APR	07:29	0	0
	391	1012	222	2	APR	08:01	0	1
X	390	1012	222	2	APR	08:45	0	0
X	391	1012	222	2	APR	09:18	0	0
*	390	1012	222	2	APR	09:49	0	0
X	320	1012	222	2	APR	11:07	0	0
*	320	1012	222	2	APR	11:31	0	0
X	321	1012	222	2	APR	11:31	0	0
X	321	1012	222	2	APR	11:43	0	0
X	320	1012	222	2	APR	11:57	0	0
X	321	1012	222	2	APR	12:18	0	0
X	320	1132	222	2	APR	12:46	0	0
*	321	1132	222	2	APR	13:16	3	0

by adding the ridership of all routes in the garage to the unknown ridership for the garage for the day of interest. An average fare is defined and used instead. Pass riders are not included since they are already estimated through one of the methods described in Section 3.4. They can simply be added to the non-pass passengers estimated in this section.

One reason for the unknown ridership at the garage level to be unreliable is that, unlike fare revenue and pass swipes, the ridership data depend heavily on the operators pressing the right buttons when a fare is deposited. As discussed earlier, if two student riders insert 30 cents each into the electronic farebox, the farebox records 60 cents of fare revenue no matter what the operator presses (except any special button such as the bypass knob). This 60 cents is going to be counted as the correct fare revenue, whether as classified fare revenue for the signcode, unclassified fare revenue for the signcode, unknown fare revenue for the garage, or even fare revenue for another route. If the operator presses the student fare button twice, then the two students are going to be counted as two passengers. However, if the operator presses the dump button, the two students are not going to be counted at all, or if the operator does not press any button, the two students are going to be counted as one regular adult passenger. Of course, there are many other possibilities depending on which buttons the operator presses. Therefore, we can see that the ridership data will generally be less accurate than revenue data.

In order to analyze ridership, another method which avoids unknown ridership is proposed. This method depends on the definition of an average fare per route, and the estimated daily route fare revenue. Average fare is the average amount of cash paid by all non-pass passengers on a route including special fare passengers such as free riders, students, and senior citizens. All revenue and ridership are aggregated at the piece of work level to generate an average fare. A histogram is then produced for each route showing the distribution (in ascending average fare) of different kinds of information:

1. trip
2. piece of work
3. revenue
4. fare-paying ridership

Table 3.9 shows the histogram for route 39 based on 1996 data. The calculated median is based on pieces of work. The four columns at the right show the distribution (in ascending average fare calculated at the piece of work level) of the pieces of work, revenue, trips, and fare-paying ridership, respectively. For example, the table shows that 5 pieces of work have an average fare of less than 10 cents. These 5 pieces of work have a total revenue of $7.02, include 5 trips, and have 135 fare riders.

Furthermore, the average fare can also be obtained by simply dividing the total fare revenue collected by the total fare-paying ridership of the route. Table 3.10 compares the different average fares calculated based on 1996 data. The first column shows the route numbers, the second through fifth columns show the average fares calculated based on median piece of work, median fare revenue, median trip, and median fare-paying ridership in the winter, respectively. The sixth column shows the average fares calculated simply by dividing the total fare revenue collected by the total fare-paying ridership in the winter. Columns 7 through 11 contain essentially the same kind of information for the spring.

In the analyses in this thesis, average fares are calculated based on median pieces of work. The reason is that each piece of work represents a single driver, and the purpose of this study is to estimate average fares that can overcome the effects of driver farebox entry errors; therefore, it is reasonable to count each driver (piece of work) as a sample observation. Furthermore, the median is used instead of the mean to avoid bias introduced by outliers, in this case, drivers who generate extremely low or high average fares. This choice is supported by Table 3.10 too. Although the average fares estimated by different methods are quite similar to each other, the average fares based on pieces of work appear to be better than the others. In the winter of 1996, the average fares based on pieces of work for 14 out of 25 routes are either a statistical mode if a mode exists, or a mode does not exist for the route. In the spring of 1996, the average fares based on pieces of work for 18 out of 25 routes have such characteristics. Table 3.11 shows the average fares based on pieces of work by route for 1994 (pooled together for winter and spring), 1996 (pooled together for winter and spring), and spring 1996. The spring 1996 average fares are the ones that are actually used in the tests for all four seasons analyzed in this thesis.

Table 3.9: Average Fare Histogram for Route 39

Histogram for Route 39

Median = 0.37	Pieces of Work	Revenue	Trips	Fare Ridership
Average Fare < 10 cents	5	7.02	5	135
Average Fare 10 cents to 19 cents	19	572.20	76	3850
Average Fare 20 cents to 29 cents	99	2781.50	342	10452
Average Fare 30 cents to 39 cents	233	7455.32	913	21588
Average Fare 40 cents to 49 cents	133	4253.87	501	9636
Average Fare 50 cents to 59 cents	69	1784.92	219	3321
Average Fare 60 cents to 69 cents	33	799.84	94	1216
Average Fare 70 cents to 79 cents	3	24.34	5	33
Average Fare 80 cents to 89 cents	0	0.00	0	0
Average Fare 90 cents to 99 cents	0	0.00	0	0
Average Fare >= $1.00	4	47.23	10	33

Table 3.10: Average Fares Analysis

Rt	W_Wk	W_Rv	W_Tp	W_Rd	W_Mn	S_Wk	S_Rv	S_Tp	S_Rd	S_Mn
14	0.38	0.38	0.37	0.37	0.37	0.28	0.28	0.28	0.24	0.26
24	0.29	0.26	0.26	0.22	0.24	0.23	0.20	0.22	0.19	0.23
26	0.31	0.63	0.63	0.63	0.57	0.27	0.28	0.27	0.25	0.27
27	0.28	0.37	0.36	0.36	0.32	0.27	0.23	0.23	0.22	0.25
28	0.37	0.36	0.35	0.26	0.29	0.31	0.31	0.31	0.29	0.29
29	0.45	0.44	0.45	0.40	0.43	0.34	0.34	0.33	0.31	0.32
30	0.39	0.28	0.29	0.18	0.21	0.30	0.22	0.26	0.18	0.21
31	0.38	0.34	0.34	0.30	0.35	0.29	0.28	0.28	0.25	0.26
32	0.37	0.34	0.35	0.29	0.30	0.30	0.30	0.30	0.27	0.27
33	0.44	0.59	0.29	0.28	0.35	0.25	0.28	0.27	0.25	0.26
34	0.38	0.35	0.35	0.32	0.31	0.35	0.34	0.33	0.30	0.28
34E	0.48	0.51	0.49	0.48	0.49	0.34	0.33	0.33	0.31	0.33
35	0.35	0.35	0.34	0.34	0.35	0.30	0.30	0.30	0.27	0.27
36	0.45	0.34	0.34	0.30	0.33	0.36	0.31	0.34	0.28	0.28
37	0.40	0.40	0.44	0.29	0.33	0.26	0.30	0.28	0.23	0.26
38	0.37	0.37	0.37	0.34	0.35	0.37	0.36	0.36	0.34	0.32
39	0.43	0.43	0.43	0.40	0.41	0.34	0.34	0.34	0.32	0.32
40	0.28	0.30	0.29	0.27	0.27	0.28	0.28	0.27	0.25	0.26
41	0.41	0.42	0.41	0.41	0.43	0.35	0.35	0.36	0.34	0.34
42	0.27	0.32	0.32	0.25	0.26	0.29	0.36	0.35	0.24	0.28
46	0.44	0.43	0.44	0.40	0.42	0.33	0.34	0.34	0.34	0.33
48	0.00	0.00	0.00	0.00	0.00	0.30	0.33	0.33	0.27	0.32
50	0.41	0.39	0.39	0.35	0.36	0.31	0.31	0.31	0.27	0.28
51	0.43	0.42	0.42	0.37	0.37	0.40	0.40	0.40	0.37	0.37
52	0.42	0.40	0.38	0.25	0.27	0.30	0.29	0.30	0.28	0.30

Table 3.11: Average Fares of Test Data for 1994 and 1996

Route	Pooled 1994	Pooled 1996	Spring 1996
14	$0.53	$0.33	0.28
24	$0.41	$0.27	0.23
26	$0.43	$0.27	0.27
27	$0.42	$0.28	0.27
28	$0.45	$0.33	0.31
29	$0.40	$0.36	0.34
30	$0.49	$0.30	0.30
31	$0.50	$0.33	0.29
32	$0.48	$0.32	0.30
33	$0.61	$0.28	0.25
34	$0.49	$0.35	0.35
34E	$0.54	$0.38	0.34
35	$0.42	$0.32	0.30
36	$0.46	$0.40	0.36
37	$0.40	$0.29	0.26
38	$0.46	$0.37	0.37
39	$0.46	$0.37	0.34
40	$0.43	$0.28	0.28
41	$0.42	$0.36	0.35
42	$0.42	$0.29	0.29
46	$0.41	$0.37	0.33
48	$0.24	$0.30	0.30
50	$0.42	$0.34	0.31
51	$0.49	$0.41	0.40
52	$0.47	$0.33	0.30

From Table 3.11, we can see that most routes have a lower average fare in 1996 than in 1994, and the spring 1996 ones are even lower. This is reasonable because driver compliance has been improving since 1994, and the improvement was especially dramatic in spring 1996. Better driver compliance means more riders are recorded by the farebox, which means that the fare per passenger (average fare) is lower. As mentioned before, poor driver compliance may cause some fare revenue to be recorded as unknown or others, but this fare revenue is still in the farebox. On the other hand, pass swipes and ridership could be lost forever if drivers do not record them into the farebox by pressing the appropriate buttons. Therefore, poor driver compliance could cause a lower recorded ridership with the fare revenue being unchanged, which generates a higher average fare. Since it is apparent that the spring 1996 average fares are more accurate than the other ones, and there is no reason to think that average fares changed significantly between 1994 and 1996, all ridership tests in this thesis use the spring 1996 average fares. However, it is still possible for spring 1996 average fares to be overestimated.

This method inherits the hypotheses of Method 3 for estimating daily fare revenue by route (see Section 3.3). All data needed for this analysis can be found in the Trip-Level Revenue and Ridership Summary and the Revenue and Ridership Summary by Route. This method consists of five steps:

1. Apply the screening procedure of Method 3 for estimating daily route fare revenue. However, since this part of the analysis of ridership involves both fare revenue and pass swipes, a couple of points need to be noted here. First, in the third and fourth rounds of screening, the record is eliminated if either the cash-registering or the pass-registering mechanism fails. Second, in the second round of screening, fare revenue, pass swipes, and ridership of two trips are all accumulated if the two trips are less than half the median trip time of that route apart.

2. Calculate the average fare for each route based on the data that have passed the screening procedure. This average fare is calculated at the level of piece of work. For this analysis, all trips on a route with the same run numbers on the same day are considered to be part of the same piece of work. For a piece of work, the total number of riders, the total number of pass swipes, and the total fare revenue are calculated. The average fare for this

piece of work is calculated by dividing the total fare revenue by the total number of fare-paying riders: the total number of riders minus the total number of pass swipes.The average fare for route j can be obtained by simply selecting the median of all the average fares calculated for different pieces of work across all days. (Again, note that in the tests in Chapter 4, only spring 1996 average fares are used.)

3. Calculate the fare revenue for each route for the day of interest. This is the allocation part of Method 3 for estimating daily route fare revenue (since the screening is performed in Step 1 above). This step allocates the unknown fare revenue of the garage to all routes for the day of interest.

4. Calculate the number of pass riders (pass swipes) for each route for the day of interest by applying Method 3 (see Section 3.4.3).

5. Calculate the number of riders for each route for the day of interest by summing the number of fare-paying riders and the number of pass riders for each route. The number of fare-paying riders for a route is obtained by dividing the fare revenue for the route (from step 3) by the average fare for that route (from Step 2).

3.6 Route Ridership by Time Period

The estimation of route ridership by time period is important because besides being of direct use; the results can be used, with appropriate conversion factors, to estimate the maximum passenger load by route and time period. This information is critical for service planning and vehicle scheduling. In this research, each day is divided into 5 time periods:

1. Early Morning (Midnight - 7 AM)
2. AM Peak (7 AM - 9 AM)
3. Base (9 AM - 4 PM)
4. PM Peak (4 PM - 6 PM)
5. Evening (6 PM - Midnight)

The proposed method consists of two parts. The first part involves obtaining the ridership of each route for the day of interest by applying the method proposed in Section 3.5. The second part involves allocating the ridership to the five time periods of the day with the daily ridership serving as the control total. The allocation part hypothesizes that there are two major sources of errors and handles them accordingly. The first source of errors is missing trip records. The second source of errors is the undercounting of fare riders due to factors such as transfers, back-door boardings, and incorrect farebox entries by the drivers. The method handles the first source of errors by adjusting the recorded ridership to account for the missing trips. The second source of errors is handled by allocating the "extra" ridership (from the daily route ridership estimation that is not accounted for by the first source of error) to all five time periods proportional to the percentage of daily fare ridership each time period has. In addition, this method is also based on the same hypotheses as the method for estimating ridership by route and day. The advantage of this method is that it does not involve the allocation of unknown ridership at the daily or trip level, which, as explained in Section 3.5, is not very accurate.

Table 3.12 shows the number of scheduled trips for each time period for each route in Bartlett Garage, which is the garage where the testing in chapter 4 is based.

This method consists of five steps:

1. Estimate the route ridership for the day of interest by applying the method proposed in Section 3.5.
2. Adjust the recorded ridership for each time period of each route to account for the missing trip records. This is performed by multiplying the recorded ridership by the number of scheduled trips divided by the number of recorded trips.
3. Calculate the amount of "extra" ridership (due to the second source of error) that should be allocated to each time period of each route by simply subtracting the sum of the adjusted ridership from Step 2 above for all five time periods for each route from the estimated daily route ridership from Step 1.
4. Allocate the "extra" ridership (from Step 3) to each time period of each route. For each route, each time period should be allocated a fraction of the "extra" ridership equal to the fraction of fare riders that time period has out of the whole day. That is, multiply the "extra" ridership (from Step 3) by the number of fare riders for the time period divided by the total number of fare riders for the day for the route.
5. The estimated ridership of each time period is calculated by adding the adjusted recorded ridership (from Step 2) to the allocated "extra" ridership (from Step 4).

When a time period has no valid data, the ridership for all time periods cannot be estimated. For routes with no valid data for a time period for some days, the time period ridership for these days is simply estimated as the mean time period ridership for those days with data for that time period.

Table 3.12: Number of Scheduled Trips for Each Time Period for Each Route in Bartlett Garage

Route	Early Morning	AM Peak	Base	PM Peak	Evening
14	3	4	14	4	2
24	10	12	23	8	18
26	10	8	19	8	13
27	7	8	16	8	5
28	27	38	77	38	55
29	12	15	18	15	29
30	12	12	36	11	17
31	28	30	74	30	42
32	37	54	71	38	41
33	3	4	14	4	3
34	33	23	33	24	25
34E	12	14	28	11	17
35	10	13	35	16	10
36	33	24	33	16	30
37	9	13	30	16	12
38	9	10	26	10	11
39	44	71	134	54	98
40	3	8	20	8	5
41	15	14	29	12	17
42	13	23	50	12	21
46	1	8	28	8	6
48	0	0	14	1	0
50	3	10	24	12	7
51	6	11	23	12	12
52	2	10	22	7	6

Chapter 4

The MBTA Case Study

This chapter describes the analyses used to investigate the methods introduced in Chapter 3 and presents the results of these analyses along with the conclusions. We begin with a summary of the objectives of the analyses and a description of the test data, then describe the sampling structure, and finally present and discuss the results of the analyses for the proposed methods for the following four types of information:

1. Daily Fare Revenue by Route
2. Daily Pass Swipes by Route
3. Daily Ridership by Route
4. Ridership by Route and Time Period

4.1 Statistical Aspects

The test data for the proposed methods are for routes from a single garage of the Massachusetts Bay Transportation Authority bus system and are drawn from the winters and springs of 1994 and 1996. The reason for analyzing only a single garage is simplicity. Two years of data are analyzed to see the effect on the methods and the overall utility of farebox data of very different levels of operator compliance. The 1994 data was selected because it was in the early days of EF technology usage at the MBTA, and driver compliance was spotty. In 1996 the MBTA was beginning a serious initiative to use EF data for route performance monitoring, and a major effort was being made to increase driver compliance and improve correct usage of the farebox. Thus the results for 1996 might be more representative of what is achievable in the future at the MBTA as well as at other transit agencies with well established EF technology. Bartlett Garage was selected as the garage for analysis for two reasons. First, Bartlett is a large MBTA garage which operates 25 routes (see Table 4.1). If the analyses for the routes in this garage are positive, then similar methods are likely to be equally effective at other MBTA garages. Second, Bartlett does not have any routes that are operated out of another garage during certain time periods (see note on Table 4.1). This avoids needless complexity in the analysis.

One point to note is that routes 34 and 34E are analyzed separately in Methods 3 and 4 instead of as a single route as in Methods 1 and 2. The reason is that Methods 3 and 4 use the median inter-record time concept, and the median inter-record time for route 34 is very different from that for route 34E.

Table 4.1: MBTA Bus Garages

Garage	Number of Bus Routes
Albany	17
Bartlett (Arborway)	25
Cabot	33
Charlestown	30
Fellsway	20
Lynn	26
North Cambridge	3 (Trackless Trolleys)
Quincy	19
Somerville	21

Note: Some routes are operated out of one garage during certain time periods and out of another during other time periods. For example, Route 60 (Kenmore Station - Chestnut Hill) is operated out of Albany Garage during A.M. Peak, Base, and P.M. Peak on Weekdays, but it is operated out of Cabot Garage during Late Night on Weekdays and on Saturdays and Sundays.

4.1.1 Sampling Structure

The testing procedure is conceptually straightforward. The basic strategy is to estimate the characteristic of interest for the winter and the spring timetables and see if significant differences exist for each route in the garage. This strategy is used because of its simplicity and because the characteristics are not expected to vary significantly between the winter and the spring of the same year for most routes. This expectation is confirmed by the garage level statistics for the two timetables. However, for pass swipe analyses, the strategy is slightly modified because some pass swipe control totals are different between the winter and the spring of the year. Instead of estimated pass swipes, estimated pass swipe shares are compared. That is, the shares of garage-level pass swipes of each route of winter and spring are compared. In addition, for the route ridership by day and time period analyses, ride check data for routes 39 and 42 for winter 1996 are used to compare with the test results.

Each characteristic is estimated for each method based on a sample of weekdays of data. The number of days used for each method depends on data availability and the complexity of the method as will be described in the following subsections. With respect to the sample itself, days were excluded for two reasons: either no data were available (typically because the data could not be distinguished from those of the following day because probing was not performed every day), or the data were clearly atypical for one of the following reasons:

1. Days that are close to a holiday are likely to have atypical data because many people take these days off. Examples are the Friday following Thanksgiving and the weekdays between Christmas and New Year's Day.
2. Days having severe weather conditions generally have atypical data because many schools and workplaces do not operate normally on these days. This affects the travel pattern of commuters which, in turn, affects the revenue and ridership data.

Usually similar sample sizes were selected for the winter and the spring of the same year. The reason for structuring the tests by timetable is that drivers usually select their runs by timetable, and the same driver is likely to be operating the same run for the whole timetable. Since the tests are designed to test the stability of the estimated revenue and ridership information, and driver habits are unlikely to change within a timetable, significant changes in driver compliance at the route level are most likely to occur at the change of timetables. Thus by estimating route level characteristics for separate timetables we can see how variable levels of driver compliance affect the forecasts using each method.

Essentially similar tests were run for 1994 and 1996 to explore the effects of different levels of operator compliance. Compliance improved significantly in 1996 after a series of initiatives including giving each operator an electronic farebox quick reference guide (see Appendix B). Hence the testing of the two sets of data gives us some idea of the effects of improved driver compliance on data quality and the effectiveness of each method under different conditions.

Since Methods 1 and 2 for estimating the items of interest at the route level do not deal directly with trip-level data, they are simpler than Methods 3 and 4. Therefore, a sample of 40 days from each season for both 1994 and 1996 was used to test Methods 1 and 2. After screening out days with bad (or no) data, this represents essentially a full sample of weekdays for a single timetable. For Methods 3 and 4 however, considerable processing was required on each day, so it was necessary in light of time and resource constraints to use a much smaller sample size to test these methods. Specifically a sample size of 5 days (5 days in each winter and 5 days in each spring) was used for each year for Methods 3 and 4. To minimize sampling bias, the sample was typically specified as the first 5 days with valid data in each season in 1994. However, this scheme was not followed for 1996 data because there were some unusual factors affecting the first few days in winter. In addition, since farebox probing was not performed daily in the winter and the early spring of 1996 (it was performed every other day), the control totals for a probe day are divided equally between this day and the previous day, when probing was not performed. The trip-level data are not affected because they are organized by transit date instead of probing date.

For the ridership estimates only the five-day sample was used. The reason for using a five-day sample is that the ridership analysis is considerably more complex than fare revenue and pass swipe analyses, and using a 40 day sample would create unnecessary complexity in the analysis. Furthermore, since the ridership estimates make use of the fare revenue estimates, the sample used is the same as that used for the fare revenue analysis.

Table 4.2 shows the days included in the sample for Methods 1 and 2. Table 4.3 shows the sampled days for application of Methods 3 and 4. In the winters of 1994 and 1996, the 5-day sample for the fare revenue analysis is different from that for the pass swipe analysis: the days used for the fare revenue and ridership analyses are without parentheses, whereas those used for the pass swipe analysis are in parentheses. Different days were used for the two analyses in order to make sure that the fare revenue and pass swipe control totals are not statistically different between the winter and the spring of 1996. As mentioned before, only three methods are proposed for pass swipe analyses. Table 4.4 shows the total days of week of the sample for Methods 1 and 2, and for the pass swipe analysis (in parentheses). Table 4.5 shows the same information for Methods 3 and 4. See Appendix C for calendars for the four seasons (winters and springs of 1994 and 1996) with the reasons for eliminating specific days.

Table 4.2: Days analyzed in Methods 1 and 2[a]

Winter 1994	Spring 1994	Winter 1996	Spring 1996
1/5 W	4/12 T	1/3 W	3/26 T
1/6 R	4/13 W	1/4 R	3/27 W
1/7 F	4/14 R	1/5 F	3/28 R
1/12 W	4/20 W	1/9 T	3/29 F
1/13 R	4/21 R	1/10 W	4/2 T
1/19 W	4/22 F	1/11 R	4/3 W
1/20 R	4/26 T	1/12 F	4/4 R
1/21 F	4/27 W	1/17 W	4/5 F
1/25 T	4/28 R	1/18 R	4/9 T
1/26 W	4/29 F	1/19 F	4/10 W
1/27 R	5/3 T	1/23 T	4/11 R
1/28 F	5/4 W	1/24 W	4/12 F
2/1 T	5/5 R	1/25 R	4/17 W
2/2 W	5/6 F	1/26 F	4/18 R
2/3 R	5/10 T	1/30 T	4/19 F
2/4 F	5/11 W	1/31 W	4/23 T
2/8 T	5/12 R	2/1 R	4/24 W
2/10 R	5/13 F	2/2 F	4/25 R
2/11 F	5/17 T	2/6 T	4/26 F
2/15 T	5/18 W	2/7 W	5/1 W
2/16 W	5/19 R	2/8 R	5/2 R
2/17 R	5/20 F	2/9 F	5/3 F
2/23 W	5/24 T	2/22 R	5/7 T
2/24 R	5/25 W	2/23 F	5/8 W
2/25 F	5/26 R	2/27 T	5/9 R
3/1 T	6/1 W	2/28 W	5/10 F

Table 4.2: Days analyzed in Methods 1 and 2[a]

Winter 1994	Spring 1994	Winter 1996	Spring 1996
3/2 W	6/2 R	2/29 R	5/14 T
3/4 F	6/3 F	3/1 F	5/15 W
3/8 T	6/7 T	3/5 T	5/16 R
3/9 W	6/8 W	3/6 W	5/17 F
3/10 R	6/9 R	3/7 R	5/21 T
3/11 F	6/10 F	3/8 F	5/22 W
3/15 T	6/14 T	3/12 T	5/29 W
3/16 W	6/15 W	3/13 W	5/30 R
3/17 R	6/16 R	3/14 R	5/31 F
3/18 F	6/17 F	3/15 F	6/4 T
3/22 T	6/21 T	3/19 T	6/5 W
3/23 W	6/22 W	3/20 W	6/6 R
3/24 R	6/23 R	3/21 R	6/7 F
3/25 F	6/24 F	3/22 F	6/11 T

a. T=Tuesday, W=Wednesday, R=Thursday, F=Friday

Table 4.3: Days analyzed in Methods 3 and 4[a]

Winter 1994	Spring 1994	Winter 1996	Spring 1996
1/5 W (2/4 F)[b]	4/12 T	1/26 F (3/14 R)	3/26 T
1/6 R (2/8 T)	4/13 W	1/30 T (3/15 F)	3/27 W
1/7 F (2/10 R)	4/14 R	1/31 W (3/19 T)	3/28 R
1/12 W (2/11 F)	4/20 W	2/1 R (3/20 W)	3/29 F
1/13 R (2/15 T)	4/21 R	2/2 F (3/21 R)	4/2 T

a. T=Tuesday, W=Wednesday, R=Thursday, F=Friday
b. In the winters of 1994 and 1996, days used for fare revenue and ridership analyses are without parentheses; days in parentheses are used for pass swipe analysis.

Table 4.4: Total Days of Week in Sample (Methods 1 and 2)

Day of Week	Winter 1994	Spring 1994	Winter 1996	Spring 1996
Tuesday (T)	8	9	8	9
Wednesday (W)	11	11	10	11
Thursday (R)	11	11	11	10
Friday (F)	10	9	11	10

Table 4.5: Total Days of Week in Sample (Methods 3 and 4)

Day of Week	Winter 1994	Spring 1994	Winter 1996	Spring 1996
Tuesday (T)	0 (2)[a]	1	1 (1)	2
Wednesday (W)	2 (0)	2	1 (1)	1
Thursday (R)	2 (1)	2	1 (2)	1
Friday (F)	1 (2)	0	2 (1)	1

a. In the winter of 1994, items without parentheses are the number of days of week sample for the fare revenue and ridership analyses; items in parentheses are the days of week sample for the pass swipe analysis.

4.1.2 Statistical Tests

For each method and each characteristic, the mean, standard deviation, and coefficient of variation of the data item for each route across all sampled days in the winter and (separately) the spring of the same year are estimated and compared. For pass analysis, the shares of garage-level pass swipes of each route for the winter and spring are compared. The share is obtained by dividing the estimated pass swipes for a route by the total number of pass swipes for the garage. A statistical test (z-test if the sample size of each season is greater than or equal to 30, t-test otherwise) is then applied to each route to decide if the estimated winter characteristic is statistically different from that of the spring. The percentage of routes that are classified as different under each method is then calculated: the smaller the percentage the better the method. For simplicity we will classify a method as having failed if more than 10% of the routes are classified as being different. Otherwise, it is considered to have passed this test. The 10% threshold is chosen somewhat arbitrarily, but the tests can easily be modified to use other threshold values should this need arise.

This approach to testing is based on the hypothesis that for most routes the winter characteristic should not be statistically different from that of the spring.

The z-test is performed at the 95% level of confidence (0.05 level of significance) with the null hypothesis that the mean of the characteristic for a route in the winter is the same as that in the spring of the same year. The alternative hypothesis is that they are not the same. The following equation shows how the z-value is computed:

$$z = (m1 - m2) / (\sqrt{(s1)^2/(n1) + (s2)^2/(n2)})$$

where:

z is the calculated z-value

$m1$ is the mean (estimated) characteristic for a route in the winter

$m2$ is the mean (estimated) characteristic for the route in the spring

$s1$ is the (estimated) standard deviation for the route in the winter

$s2$ is the (estimated) standard deviation for the route in the spring

$n1$ is the number of days in the winter sample

$n2$ is the number of days in the spring sample

If the absolute computed z-value is greater than or equal to 1.96, then the route is labelled "different"; otherwise, it is labelled "not different".

The z-test is used only when the sample size in each season is at least 30, since only then the central limit theorem applies, which assumes that the sample is normally distributed. If the sample size is smaller than 30, the t-test is used instead of the z-test, as described below.

The t-test is also performed at the 95% level of confidence (0.05 level of significance) with the same null and alternative hypotheses. The following equations show how the t-value is computed:

$$t = (m1 - m2) / (\sqrt{(sp)^2/(n1) + (sp)^2/(n2)})$$

where:

$$(sp)^2 = ((n1 - 1) \times (s1)^2 + (n2 - 1) \times (s2)^2) / (n1 + n2 - 2)$$

and:

t is the calculated t-value.

The critical t-values are used to determine if a route should be labelled "different" based on the sample size. For example, if the sample size is 5, then the critical values are ± 2.306.

Finally, the results of all methods for estimating each kind of information are compared.

4.2 Daily Fare Revenue

This section presents the analysis results for the four proposed methods for estimating fare revenue by route at the daily level. The results for each method are first presented independently, and in the final section the four methods are compared. Table 4.6 shows the means and standard deviations of the total fare revenue for Bartlett Garage for both the 40-day and 5-day samples for winter and spring for both 1994 and 1996. Also shown are the z-values/t-values which compare the winter and spring samples for each year, and whether each spring sample is statistically different from the corresponding winter sample. Since Methods 3 and 4 are more complex than Methods 1 and 2, data from only 10 days (5 in each period) are analyzed to reduce the amount of effort required. Therefore the statistics of the control totals of the 5-day samples are different from their 40-day counterparts. Also, since the 5-day samples have sample sizes of less than 30, the t-test, instead of the z-test, is used.

None of the four samples has statistically different control totals between the winter and the spring. This is important because otherwise it is difficult to determine if any statistical difference in estimated fare revenue between two periods at the route level is due to non-farebox-related factors. In addition, the average amount of unknown fare revenue per day (AUFR) that must be allocated to the routes for each period are also shown in the figure. The average amount of unknown fare revenue in 1996 is lower than that of 1994, especially for spring 1996, which is only about half of that of any other period. This result is consistent with the major effort to improve driver compliance, which occurred in the spring of 1996.

Table 4.6: Fare Revenue Control Totals

		1994 40-day Sample	1994 5-day Sample	1996 40-day Sample	1996 5-day Sample
Winter	Mean	13394.34	13415.10	12623.56	12991.19
	SD	1105.10	765.78	1228.32	938.57
	AUFR	1933.79	2126.00	1492.61	1704.12
	SD	271.45	428.69	331.06	262.54
Spring	Mean	13279.86	13660.64	12968.03	12916.31
	SD	1001.65	1296.30	1411.19	651.29
	AUFR	1691.80	1861.25	822.77	734.91
	SD	279.02	210.89	252.56	170.16
Z/T Value		0.49	-0.36	-1.16	0.15
Different?		No	No	No	No

SD = standard deviation
AUFR = average amount of unknown fare revenue per day

4.2.1 Method 1: Proportional Allocation of Unknown Fare Revenue

Table 4.7 shows the estimated daily fare revenue and coefficients of variation for each route for both winter and spring for 1994 and 1996 with routes ordered by declining revenue for ease of interpreting the impacts of route size. It also shows whether the means are statistically different between the winter and spring of each year (indicated by an *) and the z-value. In addition, any coefficient of variation at least 0.3 is indicated by a # sign.

A number of observations can be made from the results shown in Table 4.7. First, we can see that the coefficients of variation are generally lower for large routes (routes with mean estimated daily fare revenue greater than or equal to $1000 such as routes 28, 32, 34, and 39), typically below 0.2. Second, all large routes except route 28 are labelled not different, that is, with absolute z-values less than 1.96. This suggests that Method 1 can work reasonably well for estimating daily fare revenue for the largest routes. Unfortunately, it does not work very well for the smaller routes since overall 50% of the routes are labelled different in 1994 and 25% in 1996.

Table 4.7: Daily Fare Revenue: Method 1 Results

	1994					1996				
	Winter		Spring		Z-Val	Winter		Spring		Z-Val
Rt	Mean	COV	Mean	COV		Mean	COV	Mean	COV	
39	3260	0.10	3195	0.08	0.97	3090	0.11	3072	0.10	0.24
28	1927	0.13	2109	0.14	-3.04*	1771	0.13	2031	0.17	-4.01*
34	1343	0.12	1348	0.11	-0.14	1267	0.18	1337	0.12	-1.58
32	1172	0.18	1251	0.15	-1.80	1222	0.17	1248	0.13	-0.63
31	816	0.14	760	0.13	2.31*	754	0.20	796	0.17	-1.31
36	617	0.22	656	0.19	-1.33	601	0.21	611	0.19	-0.38
42	559	0.20	527	0.19	1.34	530	0.19	546	0.27	-0.59
35	412	0.22	418	0.14	-0.31	378	0.19	409	0.20	-1.79
51	403	0.20	420	0.14	-1.08	366	0.16	340	0.23	1.73
29	401	0.19	350	0.27	2.68*	370	0.14	325	0.18	3.54*
30	243	0.24	315	0.19	-5.56*	309	0.21	236	0.19	5.85*
37	332	0.15	270	0.23	4.95*	245	0.35#	254	0.22	-0.56
24	274	0.22	236	0.24	3.00*	265	0.20	277	0.22	-0.92
26	320	0.22	182	0.33#	9.46*	236	0.24	274	0.22	-2.86*
50	183	0.26	175	0.24	0.83	177	0.22	186	0.27	-0.87
38	204	0.21	194	0.25	0.90	154	0.29	160	0.22	-0.68
40	139	0.26	167	0.27	-3.09*	186	0.30	160	0.26	2.30*
41	150	0.30#	152	0.23	-0.24	169	0.27	169	0.26	0.06
52	162	0.20	173	0.29	-1.12	141	0.17	143	0.28	-0.31
27	167	0.33#	120	0.22	4.80*	95	0.28	98	0.31#	-0.37
46	97	0.47#	66	0.29	3.97*	93	0.25	96	0.22	-0.55
14	110	0.27	80	0.37#	4.39*	75	0.32#	84	0.39#	-1.27
33	53	0.51#	66	0.43#	-2.09*	90	0.33#	84	0.46#	0.73
48	50	0.45#	50	0.26	0.06	41	0.31#	34	0.39#	2.63*
T	13394	0.22	13280	0.23		12625	0.21	12970	0.22	

* The means are statistically different between the winter and spring

Coefficient of variation >= 0.3

T Total mean estimated fare revenue for the garage and median COV

In general, the coefficients of variation are quite high - coefficients of variation higher than 0.3 are flagged in the table. Based on default values table in the *Transit Data Collection Design Manual*[1], a value of 0.3 or higher is unlikely for daily route revenue. In winter 1994, only 5 out of 24 routes (routes 27, 33, 41, 46, and 48) have coefficients of variation higher than 0.3, and only 3 out of 24 routes (routes 14, 26, and 33) in the spring do. In 1996, 4 routes in each season (routes 14, 33, 37, and 48 in the winter, and routes 14, 27, 33, and 48 in the spring) have coefficients of variation greater than 0.3. All of these routes, except route 37, are small routes with mean estimated daily fare revenue below $200.

Overall the method is not satisfactory for estimating daily fare revenue by route, however, it does perform much better in 1996 than in 1994 with the improvement probably due to improved driver compliance.

4.2.2 Method 2: Allocation based on Missing Trip Records

Table 4.8 shows the number of missing trip records per day for each route ordered by declining number of scheduled trips per day for all four periods and the number of scheduled trips per day for these routes.

There are three points to note from Table 4.8:

1. In general, the average number of missing trip records decreases with time between the winter of 1994 (48% garage-wide) and the spring of 1996 (9% garage-wide). As expected, the average number of missing trip records for the spring of 1996 is significantly lower than that of any other period for all routes in the garage. In fact, two routes (34 and 39) even have negative missing trip records, probably due to multiple registrations for some trips by some operators.

1. United States Department of Transportation, "*Transit Data Collection Design Manual*", June 1985.

Table 4.8: Number of Missing Trip Records

Route	Scheduled Trips/Day	Missing Trip Records/Day (% of Scheduled Trips)			
		Winter 1994	Spring 1994	Winter 1996	Spring 1996
39	401	147 (37%)	92 (23%)	76 (19%)	-12 (-3%)
32	241	134 (56%)	108 (45%)	76 (32%)	18 (7%)
28	235	107 (46%)	79 (34%)	56 (24%)	20 (8%)
34	220	99 (45%)	80 (36%)	56 (25%)	-33 (-15%)
31	206	120 (58%)	103 (50%)	76 (37%)	36 (18%)
36	136	64 (47%)	46 (34%)	41 (30%)	6 (4%)
42	119	42 (35%)	49 (41%)	56 (47%)	11 (10%)
29	89	41 (46%)	33 (37%)	34 (38%)	15 (17%)
30	88	61 (69%)	31 (35%)	52 (59%)	27 (31%)
41	87	61 (70%)	25 (29%)	59 (68%)	20 (23%)
35	84	40 (48%)	28 (33%)	28 (34%)	10 (12%)
37	80	26 (33%)	47 (58%)	38 (48%)	8 (10%)
24	71	34 (48%)	13 (18%)	25 (35%)	3 (5%)
38	66	10 (15%)	29 (44%)	33 (50%)	8 (12%)
51	64	21 (33%)	23 (37%)	20 (31%)	9 (15%)
26	58	49 (85%)	50 (86%)	46 (79%)	43 (73%)
50	56	24 (43%)	21 (38%)	18 (32%)	8 (15%)
46	51	19 (37%)	27 (53%)	10 (19%)	1 (3%)
52	47	18 (38%)	26 (54%)	14 (31%)	1 (1%)
27	44	26 (59%)	21 (48%)	9 (20%)	2 (5%)
40	44	24 (55%)	22 (49%)	8 (18%)	4 (8%)
33	28	22 (78%)	14 (49%)	5 (18%)	7 (24%)
14	27	22 (83%)	21 (77%)	11 (42%)	6 (23%)
48	15	11 (72%)	9 (61%)	11 (73%)	4 (28%)
Total	2557	1222 (48%)	997 (39%)	858 (34%)	222 (9%)

2. Some routes (14, 26, and 48) consistently have a high percentage of missing trip records, and these are generally routes with relatively few scheduled trips per day. On the other hand, some routes (39 and 51) consistently have a low percentage of missing trip records, and these are generally routes with relatively more scheduled trips per day.

3. Just because a trip record exists does not mean that it is necessarily good. This explains the extra trip records for routes 34 and 39 for the spring of 1996. However, this method does not attempt to check the validity of a record.

Table 4.9 shows the estimated daily fare revenue and coefficients of variation for each route for both winter and spring for 1994 and 1996.

From the results, we can see that Method 2 generates similar results to Method 1. Again, all large routes have coefficients of variations below 0.2 in all seasons. In addition, all large routes except route 28 have an absolute z-value of less than 1.96. This means that the method works reasonably well for the large routes. However, similar to Method 1, Method 2 does not work well with small and medium routes since overall 50% of the routes in 1994 and 25% of the routes in 1996 have statistically different mean fare revenue figures. The coefficients of variation are generally quite low: only four coefficients of variation in the winter of 1994 (routes 14, 33, 46, and 48), three each in the spring of 1994 (routes 14, 26, and 33) and the spring of 1996 (routes 26, 33, and 48), and none in the winter of 1996 are above 0.3, and all of the routes, except route 26, with coefficients of variation greater than 0.3 are small routes with estimated daily fare revenue below $200.

Furthermore, we see a huge improvement in both the stability of data within a season and the performance of the method. First, only 25% of the routes have statistically different mean fare revenue figures in 1996 compared to 50% of the routes in 1994. Second, only 3 coefficients of variation are greater than 0.3 for the two periods in 1996 whereas there are 7 for the two periods in 1994. Therefore, we can conclude that Method 2

Table 4.9: Daily Fare Revenue: Method 2 Results

	1994					1996				
	Winter		Spring		Z-Val	Winter		Spring		Z-Val
Rt	Mean	COV	Mean	COV		Mean	COV	Mean	COV	
39	3043	0.09	2950	0.07	1.73	2863	0.10	2893	0.09	-0.50
28	1867	0.12	2027	0.12	-3.06*	1672	0.11	1972	0.14	-5.61*
34	1295	0.11	1313	0.10	-0.56	1204	0.17	1255	0.12	-1.30
32	1209	0.17	1270	0.12	-1.51	1189	0.14	1207	0.12	-0.54
31	852	0.13	806	0.13	1.92	760	0.16	809	0.15	-1.83
36	603	0.20	630	0.17	-1.06	586	0.17	590	0.17	-0.20
42	521	0.18	528	0.16	-0.37	574	0.17	545	0.23	1.17
26	535	0.18	396	0.36#	5.19*	434	0.29	596	0.40#	-3.75*
35	405	0.18	404	0.15	0.09	378	0.18	407	0.16	-1.94
51	371	0.18	410	0.14	-2.84*	359	0.15	342	0.18	1.27
29	394	0.17	346	0.22	2.96*	381	0.14	337	0.16	3.57*
30	290	0.22	308	0.17	-1.45	379	0.18	263	0.18	8.78*
37	307	0.15	311	0.24	-0.26	264	0.28	251	0.19	0.96
24	271	0.18	216	0.22	5.10*	266	0.14	273	0.18	-0.78
41	183	0.26	144	0.21	4.42*	242	0.22	178	0.20	6.35*
38	180	0.20	200	0.23	-2.12*	173	0.24	161	0.20	1.40
50	176	0.25	172	0.22	0.43	174	0.19	188	0.22	-1.64
40	144	0.23	178	0.25	-3.93*	173	0.24	158	0.22	1.86
52	153	0.20	193	0.29	-4.01*	139	0.17	137	0.26	0.31
27	182	0.25	135	0.27	5.07*	91	0.22	95	0.26	-0.82
14	187	0.34#	132	0.39#	4.32*	81	0.27	89	0.30	-1.47
46	92	0.46#	73	0.28	2.58*	87	0.21	92	0.20	-1.24
33	74	0.52#	79	0.44#	-0.52	86	0.27	89	0.35#	-0.52
48	61	0.42#	60	0.22	0.12	67	0.29	39	0.41#	7.02*
T	13395	0.19	13281	0.22		12622	0.18	12966	0.20	

* The means are statistically different between the winter and spring

Coefficient of variation >= 0.3

T Total mean estimated fare revenue for the garage and median COV

also works well with large routes, but not with small and medium routes. Furthermore, Method 2 works better with 1996 data than with 1994.

4.2.3 Method 3: Screen Bad Trip Records then Apply Method 2

Methods 3 and 4 are fundamentally different from Methods 1 and 2 since they are based on trip level screening.

4.2.3.1 Trip Screening Results

In order to understand this method, we need to start by looking at the elimination process in detail.

Table 4.10 summarizes the eliminations as a result of the screenings for both 1994 and 1996. The first column shows the route number. The second column shows the number of scheduled trips per day for each route. The third column shows the number of trips recorded on the original reports (the Trip-Level Revenue and Ridership Summaries by Route). The fourth column shows the number of trips eliminated due to defective electronic fareboxes. The fifth column shows the number of trips eliminated due to the inter-record times being too short. The sixth column shows the number of trips eliminated because of long inter-record times. The seventh column shows the number of trips remaining after all eliminations. Finally, the last column shows the number of remaining trips as a percentage of the number of scheduled trips. Please refer to Table 3.2 for the median inter-record times for each route in Bartlett Garage calculated from both the 1994 and 1996 data. These are used for the first and second rounds of screening.

From Table 4.10, a number of observations can be made. First, in 1994 the overall number of trip records is only slightly more than half (55%) the number of scheduled trips. This means that many operators were either not registering at all or registering less

Table 4.10: Daily Fare Revenue: A Summary of Eliminations Performed in Method 3

1994 Data:

Route	SchTrip	Trips	Defects	Short	Long	Remaining	(% of SchTrip)
39	4010	2698	11	258	910	1519	(38%)
32	2410	1173	8	103	421	641	(27%)
28	2350	1407	21	160	562	664	(28%)
31	2060	902	19	61	303	519	(25%)
34	1380	761	5	51	274	431	(31%)
36	1360	755	0	69	257	429	(32%)
42	1190	697	8	64	174	451	(38%)
29	890	482	9	52	211	210	(24%)
30	880	393	9	31	140	213	(24%)
41	870	411	27	73	75	236	(27%)
35	840	488	1	41	148	298	(35%)
34E	820	550	5	80	209	256	(31%)
37	800	428	10	23	133	262	(33%)
24	710	440	2	34	120	284	(40%)
38	660	468	0	73	123	272	(41%)
51	640	410	10	7	136	257	(40%)
26	580	93	3	12	68	10	(2%)
50	560	329	7	16	105	201	(36%)
46	510	295	0	6	58	231	(45%)
52	470	238	0	30	106	102	(22%)
27	440	211	13	6	54	138	(31%)
40	440	170	7	12	57	94	(21%)
33	280	97	0	29	41	27	(10%)
14	270	53	1	3	32	17	(6%)
48	150	55	0	4	37	14	(9%)
Total:	25570	14004	176	1298	4754	7776	(30%)
% of SchTrip:		55%	1%	5%	19%	30%	

1996 Data:

Route	SchTrip	Trips	Defects	Short	Long	Remaining	(% of SchTrip)
39	4010	3482	125	339	919	2099	(52%)
32	2410	1866	34	114	421	1297	(54%)
28	2350	1930	29	220	571	1110	(47%)
31	2060	1423	28	86	314	995	(48%)
34	1380	1050	31	62	305	652	(47%)
36	1360	1030	13	69	318	630	(46%)
42	1190	803	9	106	211	477	(40%)
29	890	653	3	49	235	366	(41%)
30	880	458	5	29	142	282	(32%)
41	870	429	0	11	107	311	(36%)
35	840	594	0	42	160	392	(47%)
34E	820	1005	0	140	411	454	(55%)
37	800	550	0	28	166	356	(45%)
24	710	528	8	37	153	330	(46%)
38	660	434	6	32	109	287	(43%)
51	640	473	6	17	98	352	(55%)
26	580	128	0	17	87	24	(4%)
50	560	436	5	13	112	306	(55%)
46	510	470	21	2	55	392	(77%)
52	470	343	0	52	80	211	(45%)
27	440	385	0	4	57	324	(74%)
40	440	376	6	12	93	265	(60%)
33	280	198	1	34	45	118	(42%)
14	270	166	0	5	31	130	(48%)
48	150	65	0	4	32	29	(19%)
Total:	25570	19275	330	1524	5232	12189	(48%)
% of SchTrip:		75%	1%	6%	20%	48%	

than the full number of trips run. This is confirmed by the number of trips eliminated because of a long inter-record time. A total of 4754 out of 14004 recorded trips (34%) are eliminated for this reason. However, there is significant improvement in 1996. A total of 19275 (75%) out of 25570 trips are recorded, and only 5232 out of 19275 recorded trips (27%) are eliminated because of long inter-record times. Although not all records represent legitimate trips, it is better to have extra records than missing records because extra records will be eliminated properly, and their recorded revenue will be accumulated to that of the appropriate trip by Round 1 of the elimination process.

Second, multiple registration is not as serious a problem as only 1298 out of 14004 recorded trips (9%) are eliminated in 1994 for this reason. This also improves in 1996 with only 1524 out of 19275 trips (8%) being eliminated for this reason.

Third, although the number of trips eliminated due to defective fareboxes is only 176 out of 14004 (1%), this step is still very important. Since a bus usually serves multiple trips on a route on a day, and a small route may only have one or two buses serving it, if one of these buses has a defective farebox, the daily revenue and ridership data can be severely affected. This problem is worse in 1996 than in 1994 with the number of trips deleted due to defective fareboxes increasing to 330 out of 19275 recorded trips (2%).

Fourth, we can see from Table 4.10 that data on routes 14, 26, 33, and 48 are quite problematic, especially in 1994 when all these routes had valid records for no more than 10% of the scheduled trips compared with the garage average of 30%. There is a real shortage of good data for these problem routes compounded by the fact that many of these routes had very few records to start with. For example, route 14 had only 53 records out of 270 scheduled trips (20%), much less than the garage average of 55%. Similarly, for routes 26, 33, and 48, the percentage of scheduled trips recorded (before the screenings) are 16%, 35%, and 37%, respectively. A particular reason for the poor performance of routes 26 and 48 is that they are loop routes. Some operators may not know how to register with the farebox properly on loop routes. Fortunately, all of these routes have improved in 1996, with the improvement on routes 14 and 33 particularly striking.

Finally, in 1994 the number of remaining trips is 7776, which is only about 56% of the records and 30% of the scheduled trips. This affects the small routes more than the larger routes because the small routes may end up having days with no data since they do not have much data to start with. Also, since small routes usually have only one or two operators for the whole day, if the operator does not use the correct farebox entry procedure, the route may not have any data, let alone good data, for the whole day. Fortunately, this situation has also improved in 1996. There are a total of 12189 trips left after all eliminations in 1996. This is 63% of the total records and 48% of the scheduled trips, a huge improvement from 1994. In addition, except for routes 26 and 48, all routes have at least 30% the scheduled trips remaining after the screening. This is extremely important for the small routes because they are now more likely to have enough good data to provide a reasonable estimate of the daily fare revenue.

Table 4.11 shows the number of trip records remaining after the screenings, summarized into three classes based on the number of scheduled trips per day. The first set contains the trip screening results for the large routes (with more than 100 scheduled trips per day), the second set contains the results for the medium-sized routes (between 50 and 100 scheduled trips per day), and the third set contains the information for the small routes (less than 50 scheduled trips per day). For each route the table shows:

1. the number of scheduled trips per day.
2. the average number of trips per day remaining after all screenings. For example, the average number of trips per day in winter 1994 for route 39 was 125. This means that there were 625 trips left after all screenings for the 5 days analyzed in winter 1994.
3. the number of days with no trip-level record remaining (after all screenings) among the 5 days analyzed.

Table 4.11: Daily Fare Revenue: Trip Record Screening Results

		1994				1996			
Number of Scheduled Trips > 100									
Route	SCH_T	W_TPS	0_DYS	S_TPS	0_DYS	W_TPS	0_DYS	S_TPS	0_DYS
39	401	125	0	179	0	160	0	276	0
32	241	48	0	80	0	98	0	163	0
28	235	58	0	75	0	94	0	129	0
31	206	47	0	58	0	72	0	130	0
34	138	37	0	49	0	49	0	82	0
36	136	36	0	50	0	51	0	77	0
42*	119	48	0	42	0	26	0	70	0
Number of Scheduled Trips between 50 and 100 inclusive									
Route	SCH_T	W_TPS	0_DYS	S_TPS	0_DYS	W_TPS	0_DYS	S_TPS	0_DYS
29	89	20	0	23	0	23	0	50	0
30#	88	12	0	31	0	16	0	41	0
41	87	16	0	34	0	11	0	51	0
35*	84	28	0	32	0	26	0	52	0
34E	82	23	0	29	0	10	0	80	0
37	80	33	0	20	0	19	0	52	0
24*	71	23	0	34	0	22	0	45	0
38*	66	32	0	22	0	17	0	40	0
51	64	30	0	22	0	25	0	45	0
26#*	58	1	1	2	2	1	4	4	0
50	56	22	0	19	0	23	0	38	0
46	51	27	0	19	0	34	0	44	0
Number of scheduled trips < 50									
Route	SCH_T	W_TPS	0_DYS	S_TPS	0_DYS	W_TPS	0_DYS	S_TPS	0_DYS
52#	47	14	0	6	0	16	0	27	0
27#	44	11	1	16	0	29	0	36	0
40	44	12	0	7	0	23	0	30	0
33#*	28	1	4	5	1	11	0	12	0
14#*	27	0	5	3	0	9	0	17	0
48#*	15	0	5	3	0	0	5	6	0
Total	2557	704	16	860	3	865	9	1597	0

Coefficient of variation >= 0.3 (or unavailable) in at least one season

* The means are statistically different between winter and spring (or T value unavailable) in at least one year

This information is shown for all four timetables analyzed.

From Table 4.11, we can see that this screening performs very well for the large routes. All routes have data for all days analyzed. For the medium-sized routes, other than the loop route 26, all routes have data for all days. The results for the small routes are not as encouraging. In winter 1994, 4 out of 6 routes have one or more days with no data remaining after the screenings. However, there was some improvement between 1994 and 1996. In winter 1996, only route 48 was left with no data, and in spring 1996, all small routes had data for all days.

4.2.3.2 Analysis Results

Table 4.12 shows the overall results for Method 3. The organization of Table 4.12 is slightly different from that of Tables 4.7 and 4.9. Instead of ordering the routes strictly by declining estimated fare revenue, the routes are classified into 3 categories (large, medium-sized, and small) as in Table 4.11 for ease of interpretation. Within each group, the routes are ordered by declining number of scheduled trips per day. For the large routes (refer to Table 4.11 for the classification of routes by size), the results are very good. First of all, no coefficient of variation among routes with more than 100 scheduled trips per day is above 0.19. Furthermore, all but one routes are labelled as not being different from winter to spring.

We can see that only two medium-sized routes (routes 26 and 30) have one or more coefficients of variation greater than 0.3. In fact, from Table 4.11, we can see that route 26 has one or more days with no trip record remaining in three out of the four seasons, and from Table 4.12, we can see that its coefficients of variation are high except in spring 1996 (0.562 in winter 1994 and 1.037 in spring 1994). This means that route 26 cannot be properly analyzed by this method. However, since only 4 out of 12 medium-sized routes are labelled different (routes 24, 26, 35, and 38), for 8 routes this method performs satisfactorily. We can see that this method performs reasonably well on medium-sized routes and that the data within each season are generally quite stable.

The performance for the small routes is not as good. A number of items are labelled NA in this table meaning that the item is not available. For example, the estimated mean fare revenue for route 14 in winter 1994 is not available. This means that all data for route 14 are eliminated by Method 3, leaving no records on which to base a revenue estimate (see Table 4.11). For some routes, such as route 33 in winter 1994, the estimated mean fare revenue exists but there is no coefficient of variation. This is because only one day among the five analyzed has records (see Table 4.11), so no standard deviation can be calculated. Since our purpose here is to test the method, and the routes with incomplete information cannot be estimated using this method, they are automatically labelled different in the last column of Table 4.12.

We can see that 4 of the 25 routes are classified NA for at least one period: routes 14, 26, 33, and 48. This means that they cannot be properly analyzed by this method with the sample data available. One similarity among them is that they are all relatively small routes, typically with less than 50 scheduled trips per day. Even route 26, the largest of these four with 58 scheduled trips per day, has on average 1 or 2 trip records remaining after the screenings. From this, we can conclude that this method is unreliable for small routes.

The only small route with satisfactory results is route 40. All other small routes have high coefficients of variation.

There is also a clear difference between the quality of the two sets of data. First, there are fewer days with no data remaining in 1996 than in 1994, which in turn leads to fewer NA labels in 1996. In fact, in the spring of 1996, there is no route with no data for any

Table 4.12: Daily Fare Revenue: Method 3 Results

	1994					1996				
	Winter		Spring		T-Val	Winter		Spring		T-Val
Rt	Mean	COV	Mean	COV		Mean	COV	Mean	COV	
39	2935	0.09	3070	0.09	-0.80	3302	0.06	3151	0.06	1.27
32	1438	0.09	1384	0.11	0.62	1280	0.11	1328	0.05	-0.68
28	2144	0.05	2146	0.11	-0.02	1832	0.12	1982	0.06	-1.41
31	866	0.07	751	0.18	1.75	792	0.08	742	0.10	1.16
34	602	0.12	666	0.17	-1.08	595	0.05	553	0.06	2.14
36	574	0.05	652	0.14	-1.76	605	0.04	599	0.07	0.27
42	524	0.19	644	0.09	-2.35*	687	0.10	528	0.10	4.15*
29	456	0.08	518	0.25	-1.05	371	0.28	337	0.08	0.71
30	225	0.32#	273	0.18	-1.24	317	0.17	277	0.14	1.35
41	164	0.25	150	0.11	0.70	251	0.27	205	0.07	1.48
35	563	0.14	358	0.25	3.82*	363	0.21	426	0.06	-1.75
34E	665	0.15	810	0.13	-2.17	475	0.20	451	0.07	0.54
37	320	0.11	298	0.19	0.71	313	0.15	267	0.16	1.58
24	304	0.19	219	0.07	3.26*	299	0.10	263	0.07	2.29
38	224	0.12	176	0.09	3.45*	148	0.19	177	0.08	-2.03
51	388	0.10	431	0.16	-1.21	389	0.18	379	0.06	0.32
26	309	0.56#	338	1.04#	-0.16	249	NA#	354	0.28	NA*
50	186	0.28	186	0.11	-0.00	162	0.05	179	0.15	-1.37
46	100	0.18	90	0.21	0.85	96	0.13	91	0.14	0.58
52	172	0.31#	154	0.10	0.73	160	0.15	144	0.20	0.96
27	141	0.32#	102	0.13	1.85	113	0.07	101	0.15	1.67
40	176	0.09	195	0.30	-0.72	173	0.11	173	0.14	-0.01
33	149	NA#	104	0.17	NA*	119	0.14	87	0.14	3.52*
14	NA	NA#	73	0.46#	NA*	98	0.11	101	0.15	-0.33
48	NA	NA#	30	0.49#	NA*	NA	NA#	21	0.61#	NA*
T	13625	0.15	13818	0.16		13189	0.13	12916	0.10	

* The means are statistically different between the winter and spring (or T value unavailable)

\# Coefficient of variation >= 0.3 (or unavailable)

T Total mean estimated fare revenue for the garage and median COV

single day analyzed, which is especially important for the small routes. Second, we can see that the coefficients of variation in 1996 are generally lower than those in 1994. Other than the ones unavailable, the only coefficient of variation greater than 0.3 is route 48 in spring 1996 (0.61). Finally, fewer routes in 1996 have statistically different means than in 1994. In fact, only 16 percents of the routes have statistically different means in 1996, including the routes with unavailable coefficients of variation. This means that 84 percent of the routes can be satisfactorily analyzed by Method 3. Although this figure is still slightly lower than the preset 90 percent threshold, it is reasonably close. This implies that the method has the potential to perform quite satisfactorily if the data quality is maintained or improved and the sample size is increased.

Therefore, we conclude that Method 3 works much better for large and medium-sized routes than for small routes. In addition, the significant improvement of driver compliance from 1994 to 1996 makes Method 3 work better with the small routes.

4.2.4 Method 4: Combination of Methods 1 and 3

From the previous section, we see that Method 3 does a reasonably good job with medium-sized routes, a very good job with large routes, but not so well with small routes. The main reason is that since the small routes do not have a lot of data to start with, after the screening part of Method 3, some routes end up with no data for one or more days because all trips for those days are eliminated. Method 4 attempts to resolve this problem by applying Method 1 to these small routes, which at least allocates some fare revenue to these routes rather than coming up with nothing. Method 3 is retained for the remaining routes.

Table 4.13: Daily Fare Revenue: Method 4 Results

Rt	1994 Winter Mean	1994 Winter COV	1994 Spring Mean	1994 Spring COV	T-Val	1996 Winter Mean	1996 Winter COV	1996 Spring Mean	1996 Spring COV	T-Val
39	2714	0.09	2950	0.08	-1.64	3172	0.05	3125	0.05	0.47
32	1311	0.08	1320	0.08	-0.13	1230	0.10	1317	0.05	-1.40
28	1967	0.05	2044	0.09	-0.87	1759	0.12	1958	0.05	-1.94
31	792	0.07	712	0.16	1.40	759	0.09	735	0.09	0.54
34	554	0.13	637	0.16	-1.48	570	0.04	547	0.06	1.26
36	528	0.05	624	0.13	-2.45*	580	0.02	592	0.07	-0.64
42	490	0.18	615	0.09	-2.67*	652	0.09	523	0.10	3.61*
29	417	0.07	491	0.23	-1.44	353	0.27	333	0.07	0.44
30	204	0.33#	261	0.17	-1.55	299	0.16	273	0.14	0.98
41	149	0.25	145	0.12	0.25	235	0.26	202	0.07	1.16
35	521	0.14	342	0.24	3.67*	347	0.20	422	0.07	-2.27
34E	613	0.16	772	0.12	-2.66*	448	0.21	451	0.07	-0.06
37	299	0.12	284	0.20	0.52	297	0.14	265	0.16	1.20
24	281	0.17	212	0.05	3.18*	286	0.08	260	0.07	1.86
38	211	0.12	168	0.08	3.35*	141	0.19	175	0.08	-2.51*
51	366	0.11	411	0.14	-1.40	373	0.16	376	0.06	-0.10
26	841	0.23	509	0.29	3.01*&	508	0.37#	444	0.24	0.65&
50	174	0.28	177	0.08	-0.16	156	0.05	177	0.15	-1.80
46	95	0.18	87	0.20	0.76	94	0.13	91	0.14	0.39
52	159	0.31#	145	0.07	0.61	153	0.15	143	0.20	0.63
27	130	0.31#	98	0.11	1.72	111	0.06	100	0.14	1.44
40	162	0.09	184	0.29	-0.92	168	0.11	172	0.14	-0.30
33	99	0.53#	181	0.34#	-2.27&	114	0.13	86	0.14	3.38*
14	229	0.42#	187	0.22	0.91&	94	0.09	100	0.15	-0.79
48	136	0.16	104	0.39#	1.55&	96	0.35#	49	0.61#	2.33*&
T	13442	0.16	13660	0.14		12995	0.13	12916	0.09	

* The means are statistically different between the winter and spring

\# Coefficient of variation >= 0.3

& Method 1, instead of Method 3, is used in the analysis for the year

T Total mean estimated fare revenue for the garage and median COV

Table 4.13 shows the results of Method 4. As in Table 4.12, the routes are classified into 3 categories. This table is similar to the ones shown in the previous sections except that routes with a t-value followed by an & are routes that are analyzed by Method 1 instead of Method 3 in the corresponding year, and the coefficients of variation and t-values of these trips are obtained from Method 1 instead of Method 3.

From Table 4.13, we can make a number of observations. First, the percentage of routes labelled different is the same as that of Method 3: 28% in 1994, and 16% in 1996.

Second, the overall number of occurrences of high coefficients of variation is lower with Method 4 than with Method 3. A total of 5 routes in the winter and 2 routes in the spring of 1994 have coefficients of variation greater than 0.3 compared with 2 routes in the winter and 1 route in the spring of 1996. This shows improvement between 1994 and 1996 as well as an improvement over Method 3.

Third, only two routes (routes 26 and 48) must be estimated with Method 1 in 1996 compared with four in 1994. In addition, from Table 4.11, all routes in the spring of 1996 have an average of 3 or more trips per day with no day without data. This means that the driver compliance rate has improved to a level that Method 1 is no longer necessary for this analysis. However, Method 1 is still used for routes 26 and 48 in 1996 because they do not have enough recorded trips in the winter.

Fourth, routes which have no data left after the screening process of Method 3 usually have a high coefficient of variation. This is reasonable because these routes have poor data, and this is why data gets eliminated by Method 3, so one can expect the coefficients of variation to be quite high when these routes are analyzed by Method 1.

Fifth, the estimated fare revenue of each route by Method 4 is slightly lower than those of Method 3, except the ones that are analyzed by Method 1. This is simply because the routes that end up with no data after the screening do not get any control total revenue allocated to them in Method 3. In other words, the fare revenue of these routes becomes part of the unknown fare revenue that is allocated to the other routes. Hence all the control totals are only allocated to routes that have data remaining after screening. On the other hand, in Method 4 every route gets some allocation of the control total, so a route generally has a smaller estimated fare revenue in Method 4 than in Method 3 except for the routes that end up with no data after the screening of Method 3. In fact, Method 3 probably provides slightly inflated revenue estimates for most routes for this reason.

Finally, as with other methods, the larger routes generally have lower coefficients of variation than the smaller ones, except for route 26. For this reason, even though most small routes have estimated fare revenue that are not considered statistically different between the winter and the spring of 1994, the method is not considered successful for these routes because the variation within a season is too great. In 1996, however, most medium-sized and large routes appear to perform very well. In fact, other than routes 33 and 48, all small routes also perform well.

From the above observations, we can conclude that Method 4 works as well as Method 3 for large and medium-sized routes, and with good driver compliance as demonstrated with the 1996 data, it has the potential to work well with small routes. In addition, while Method 3 has difficulty in providing an estimate for the small routes with no data left after screening, Method 4 overcomes this problem by using Method 1 to estimate the daily fare revenue for small routes that have very little data left after screening. However, the quality of these estimates may not be as good as those of larger routes.

4.2.5 A Comparison of the Four Methods

Since all four methods appear to work better for large routes, it is useful to compare the results of the four methods to draw some conclusions about when to use which method, or to see if one method is significantly better than the others.

Table 4.14 shows the estimated fare revenue, coefficients of variation, and t/z statistics resulting from applying the four methods in 1996. The 1994 results are not compared because the 1996 results are much more relevant in terms of the likely future effectiveness of each method. Remember that Methods 1 and 2 are based on a sample size of 40 for each season whereas Methods 3 and 4 are based on a sample size of 5 for each season. In fact, the data used in Methods 3 and 4 are a subset of those used in Methods 1 and 2.

The performance of Methods 1 and 2 is very similar. As discussed previously, both methods have the same percentage of routes with mean estimated fare revenue statistically different between the winter and the spring of 1996 (25%). Both methods perform well for large routes and perform poorly for small routes, and both methods produce stable data within a season for large routes, and unstable data for small routes. The only advantage that Method 2 has is that in general it appears to produce slightly lower coefficients of variation than Method 1.

Although Methods 3 and 4 have the same percentage of routes with mean estimated fare revenue statistically different between the winter and the spring of 1996 (16%), Method 4 performs better than Method 3 in two ways. First, Method 4 does not leave small routes without any estimated fare revenue even when there is no data for all days after screenings since it uses Method 1 to analyze these small routes. Second, Method 4 generally produces lower coefficients of variation than Method 3, but Method 4 does not necessarily produce higher t-values. This phenomenon can be explained by the fact that

Table 4.14: Daily Fare Revenue: A Comparison of the Four Methods (1994 data)

1996 Data

Number of Scheduled Trips > 100

	METHOD 1	METHOD 2	METHOD 3	METHOD 4
Route: 39				
Winter:	3090/0.11	2863/0.10	3302/0.06	3172/0.05
Spring:	3072/0.10	2893/0.09	3151/0.06	3125/0.05
% Difference:	-1	1	-5	-1
T/Z-Value:	0.24	-0.50	1.27	0.47
Route: 32				
Winter:	1222/0.17	1189/0.14	1280/0.11	1230/0.10
Spring:	1248/0.13	1207/0.12	1328/0.05	1317/0.05
% Difference:	2	2	4	7
T/Z-Value:	-0.63	-0.54	-0.68	-1.40
Route: 28				
Winter:	1771/0.13	1672/0.11	1832/0.12	1759/0.12
Spring:	2031/0.17	1972/0.14	1982/0.06	1958/0.05
% Difference:	15	18	8	11
T/Z-Value:	-4.01*	-5.61*	-1.41	-1.94
Route: 31				
Winter:	754/0.20	760/0.16	792/0.08	759/0.09
Spring:	796/0.17	809/0.15	742/0.10	735/0.09
% Difference:	6	6	-6	-3
T/Z-Value:	-1.31	-1.83	1.16	0.54
Route: 34				
Winter:	1267/0.18	1204/0.17	1070/0.10	1017/0.10
Spring:	1337/0.12	1255/0.12	1004/0.03	998/0.03
% Difference:	6	4	-6	-2
T/Z-Value:	-1.58	-1.30	1.38	0.43
Route: 36				
Winter:	601/0.21	586/0.17	605/0.04	580/0.02
Spring:	611/0.19	590/0.17	599/0.07	592/0.07
% Difference:	2	1	-1	2
T/Z-Value:	-0.38	-0.20	0.27	-0.64
Route: 42				
Winter:	530/0.19	574/0.17	687/0.10	652/0.09
Spring:	546/0.27	545/0.23	528/0.10	523/0.10
% Difference:	3	-5	-23	-20
T/Z-Value:	-0.59	1.17	4.15*	3.61*
Total Winter:	9235	8848	9568	9169
Total Spring:	9641	9271	9335	9248
% Difference:	4	5	-2	1

Number of Scheduled Trips between 50 and 100 inclusive

	METHOD 1	METHOD 2	METHOD 3	METHOD 4
Route: 29				
Winter:	370/0.14	381/0.14	371/0.28	353/0.27
Spring:	325/0.18	337/0.16	337/0.08	333/0.07
% Difference:	-12	-12	-9	-6
T/Z-Value:	3.54*	3.57*	0.71	0.44
Route: 30				

Winter:	309/0.21	379/0.18	317/0.17	299/0.16
Spring:	236/0.19	263/0.18	277/0.14	273/0.14
% Difference:	-24	-31	-13	-9
T/Z-Value:	5.85*	8.78*	1.35	0.98

Route: 41

Winter:	169/0.27	242/0.22	251/0.27	235/0.26
Spring:	169/0.26	178/0.20	205/0.07	202/0.07
% Difference:	0	-26	-18	-14
T/Z-Value:	0.06	6.35*	1.48	1.16

Route: 35

Winter:	378/0.19	378/0.18	363/0.21	347/0.20
Spring:	409/0.20	407/0.16	426/0.06	422/0.07
% Difference:	8	8	17	22
T/Z-Value:	-1.79	-1.94	-1.75	-2.27

Route: 37

Winter:	245/0.35#	264/0.28	313/0.15	297/0.14
Spring:	254/0.22	251/0.19	267/0.16	265/0.16
% Difference:	4	-5	-15	-11
T/Z-Value:	-0.56	0.96	1.58	1.20

Route: 24

Winter:	265/0.20	266/0.14	299/0.10	286/0.08
Spring:	277/0.22	273/0.18	263/0.07	260/0.07
% Difference:	5	3	-12	-9
T/Z-Value:	-0.92	-0.78	2.29	1.86

Route: 38

Winter:	154/0.29	173/0.24	148/0.19	141/0.19
Spring:	160/0.22	161/0.20	177/0.08	175/0.08
% Difference:	4	-7	20	24
T/Z-Value:	-0.68	1.40	-2.03	-2.51*

Route: 51

Winter:	366/0.16	359/0.15	389/0.18	373/0.16
Spring:	340/0.23	342/0.18	379/0.06	376/0.06
% Difference:	-7	-5	-3	1
T/Z-Value:	1.73	1.27	0.32	-0.10

Route: 26

Winter:	236/0.24	434/0.29	249/NA#	508/0.37#
Spring:	274/0.22	596/0.40#	354/0.28	444/0.24
% Difference:	16	37	42	-13
T/Z-Value:	-2.86*	-3.75*	NA*	0.65

Route: 50

Winter:	177/0.22	174/0.19	162/0.05	156/0.05
Spring:	186/0.27	188/0.22	179/0.15	177/0.15
% Difference:	5	8	10	13
T/Z-Value:	-0.87	-1.64	-1.37	-1.80

Route: 46

Winter:	93/0.25	87/0.21	96/0.13	94/0.13
Spring:	96/0.22	92/0.20	91/0.14	91/0.14

% Difference:	3	6	-5	-3
T/Z-Value:	-0.55	-1.24	0.58	0.39
Total Winter:	2762	3137	2958	3089
Total Spring:	2726	3088	2955	3018
% Difference:	-1	-3	0	-2

Number of Scheduled Trips < 50

	METHOD 1	METHOD 2	METHOD 3	METHOD 4
Route: 52				
Winter:	141/0.17	139/0.17	160/0.15	153/0.15
Spring:	143/0.28	137/0.26	144/0.20	143/0.20
% Difference:	1	-1	-10	-7
T/Z-Value:	-0.31	0.31	0.96	0.63
Route: 27				
Winter:	95/0.28	91/0.22	113/0.07	111/0.06
Spring:	98/0.31#	95/0.26	101/0.15	100/0.14
% Difference:	3	4	-11	-10
T/Z-Value:	-0.37	-0.82	1.67	1.44
Route: 40				
Winter:	186/0.30	173/0.24	173/0.11	168/0.11
Spring:	160/0.26	158/0.22	173/0.14	172/0.14
% Difference:	-14	-9	0	2
T/Z-Value:	2.30*	1.86	-0.01	-0.30
Route: 33				
Winter:	90/0.33#	86/0.27	119/0.14	114/0.13
Spring:	84/0.46#	89/0.35#	87/0.14	86/0.14
% Difference:	-7	3	-27	-25
T/Z-Value:	0.73	-0.52	3.52*	3.38*
Route: 14				
Winter:	75/0.32#	81/0.27	98/0.11	94/0.09
Spring:	84/0.39#	89/0.30	101/0.15	100/0.15
% Difference:	12	10	3	6
T/Z-Value:	-1.27	-1.47	-0.33	-0.79
Route: 48				
Winter:	41/0.31#	67/0.29	NA/NA#	96/0.35#
Spring:	34/0.39#	39/0.41#	21/0.61#	49/0.61#
% Difference:	-17	-42	NA	-49
T/Z-Value:	2.63*	7.02*	NA*	2.33*
Total Winter:	628	637	NA	736
Total Spring:	603	607	627	650
% Difference:	-4	-5	NA	-12

Note: Each entry in the Winter row or the Spring row contains two items: estimated fare revenue and coefficient of variation, respectively.

Coefficient of variation (COV) at least 0.3 (rounding effect is considered) or unavailable
* The means are statistically different between the winter and spring of the year (or T value unavailable)

Method 3 allocates no control total revenue to routes that have no data left after screenings, which is not realistic, while Method 4 does not have this problem. Therefore, although Method 4 uses Method 3 to analyze routes with a reasonable amount of data left after all eliminations, the amount of control total (and unknown) fare revenue to be allocated to these routes is different, and the amount used by Method 4 is more realistic. However, in general, both methods perform well with medium-sized and large routes, and produce stable data within a season for these routes.

Although Methods 1 and 2 cannot be compared rigorously with Methods 3 and 4 because of the difference in sample sizes, some rough comparisons can be made. Since Method 4 clearly performs better than Method 3, only Method 4 is compared with Methods 1 and 2 here. A number of observations are made. First, Method 4 generally produces smaller coefficients of variation than Methods 1 and 2, but it does not necessarily produce higher t-values than the z-values produced by Methods 1 and 2. Second, although all methods perform better with large routes than with the smaller ones, Method 4 works better with medium-sized routes than Methods 1 and 2. In 1996, a total of 9 out of 11 medium-sized routes can be successfully analyzed by Method 4 whereas only 7 can be analyzed by Method 1 or 2. A route that can be successfully analyzed by a method is defined as a route that has coefficients of variation of less than 0.3 for both seasons and have mean estimated fare revenue not statistically different between the winter and the spring of the year. However, remember that we are being conservative in this analysis. That is, we would rather reject a good method than to accept a bad method. The very small sample size for Methods 3 and 4 implies that we can place very little confidence in the coefficient of variation figures, and it is possible that the coefficients of variation can be reduced greatly by increasing the sample size. We use 0.3 as the cutoff here simply to reduce unnecessary complexity, but the implications of the very small sample size should not be overlooked. Finally, as shown in Tables 4.7, 4.9, and 4.13, Method 4 produces a smaller percentage of route with mean estimated fare revenue significantly different between the winter and the spring of 1994 and 1996 than Methods 1 and 2. Therefore, Method 4 is the best method for estimating daily fare revenue by route.

4.3 Daily Pass Swipes

This section presents the analysis results for the three proposed methods for estimating daily pass swipes by route. Similar to the fare revenue analysis, the results for each method are presented independently, and in the final section the three methods are compared.

Table 4.15 shows the statistics on the control totals showing the means and standard deviations of the total number of pass swipes for Bartlett Garage for both the 40-day and 5-day samples for winter and spring for both 1994 and 1996. Also shown are the z-values/ t-values which compare the winter and spring samples for each year, and whether each spring sample is statistically different from the corresponding winter sample is also shown.

Unfortunately some sets of control totals have means that are statistically different between the winter and the spring of the year. This is especially true for the 40-day samples for both types of passes in 1996. Since the number of weekdays with good data in each timetable is only about 40, it is difficult to select a sample with control total means that are not statistically different between the winter and the spring of the year for both types of passes. In fact, the number of pass swipes at the garage level for both types of passes studied is probably quite different between the winter and the spring of 1996 due to improvement in driver compliance in the spring of 1996. However, with more flexibility, the 5-day samples are carefully selected so that the control total means are not statistically different between the winter and the spring of 1996.

In order to draw useful conclusions out of the analyses where the control totals are statistically different between the winter and the spring of the year, the pass swipe shares represented by each route are also analyzed. Since pass swipe shares are not affected by statistically different control totals, pass swipe share analyses supplement the total pass swipe analyses in these cases. Most of the discussion below will focus on estimated pass swipe shares rather than the estimated pass swipes themselves.

In addition, the average amount of unknown pass swipes for each pass type per day that needs to be allocated to the routes for each period are also included in the figure. The average amount of unknown pass swipes for each pass type in 1996 is lower than that of 1994, and in particular the spring 1996 numbers are only about half that of any other period. This result is consistent with the improvement in driver compliance, which occurred in the spring of 1996.

4.3.1 Method 1: Proportional Allocation of Unknown Pass Swipes

Table 4.16 shows the estimated local bus pass swipes and coefficients of variation for each route for both winter and spring of 1994 and 1996. It also shows whether the means are statistically different between the winter and spring of each year (indicated by an *) and the z-value. In addition, any coefficient of variation greater than 0.3 is indicated by a # sign. Furthermore, since the control totals are different between the winter and the spring of both 1994 and 1996, the mean local bus pass swipe shares of each route are also shown on the following row (designated by an S), together with the corresponding coefficients of variation and the z-value. Any coefficient of variation greater than 0.3 is indicated by a % sign, and if the mean shares are statistically different between the winter and the spring of the year, the route will be indicated by an X. In this table routes are ordered by declining local bus pass swipes for ease of interpreting the impacts of route size.

Table 4.15: Pass Swipes Control Totals

Local Bus Pass

		1994 40-day Sample	1994 5-day Sample	1996 40-day Sample	1996 5-day Sample
Winter	Mean	12830.58	12682.60	12584.45	14778.00
	SD	1349.66	1581.62	1593.69	1347.64
	AUFR	1797.08	1760.00	1413.00	1519.80
	SD	377.75	305.30	311.50	52.35
Spring	Mean	13801.12	14704.00	14302.88	14367.60
	SD	1081.44	1133.69	1465.73	474.77
	AUFR	1675.05	1977.20	947.45	812.00
	SD	1576.42	1426.78	306.86	109.91
Z/T Value		-3.55	-2.32	-5.02	0.64
Different?		Yes	Yes	Yes	No

Combo Pass

		1994 40-day Sample	1994 5-day Sample	1996 40-day Sample	1996 5-day Sample
Winter	Mean	8721.65	8746.80	8042.82	9272.60
	SD	745.54	918.28	968.73	842.26
	AUFR	1269.67	1292.80	1015.88	1043.60
	SD	216.48	140.08	216.07	37.77
Spring	Mean	8940.73	9682.40	8946.92	9041.00
	SD	791.21	595.28	895.58	380.55
	AUFR	1257.13	1386.60	641.50	556.40
	SD	258.49	178.53	185.36	99.00
Z/T Value		-1.27	-1.91	-4.33	0.56
Different?		No	No	Yes	No

Since the control totals are statistically different between the winter and the spring of both 1994 and 1996, the analysis of the estimated local bus pass swipe shares are more useful than that of the estimated local bus pass swipes themselves, and therefore, the discussion will focus on the shares. Table 4.16 shows that four routes in winter 1994 and three in the other three seasons have coefficients of variation of greater than 0.3, which means that there is not much improvement between 1994 and 1996. This is also confirmed by the median coefficients of variation of the estimated local bus pass swipe shares for the

Table 4.16: Daily Local Bus Pass Swipes: Method 1 Results

Rt	1994 Winter Mean	COV	1994 Spring Mean	COV	Z-Val	1996 Winter Mean	COV	1996 Spring Mean	COV	Z-Val
39	3856	0.12	4098	0.10	-2.54*	3620	0.11	4053	0.08	-5.23*
S	30.1	0.06	29.7	0.05	1.02	28.9	0.07	28.4	0.06	1.05
28	2059	0.14	2485	0.11	-6.86*	2079	0.18	2571	0.14	-6.03*
S	16.0	0.08	18.0	0.09	-6.15X	16.4	0.07	17.9	0.07	-5.73X
32	912	0.17	1024	0.15	-3.29*	1028	0.18	1128	0.12	-2.79*
S	7.1	0.16	7.4	0.12	-1.17	8.1	0.10	7.9	0.10	1.25
34	825	0.14	876	0.12	-2.10*	949	0.17	1016	0.12	-2.16*
S	6.4	0.10	6.4	0.09	0.65	7.6	0.14	7.1	0.09	2.27X
31	800	0.15	843	0.13	-1.68	758	0.28	926	0.17	-4.05*
S	6.2	0.11	6.1	0.11	0.85	6.0	0.17	6.5	0.12	-2.51X
42	770	0.21	821	0.18	-1.47	712	0.22	836	0.27	-2.83*
S	6.0	0.15	5.9	0.15	0.13	5.6	0.17	5.8	0.23	-0.63
36	440	0.27	499	0.19	-2.48*	428	0.20	487	0.19	-2.95*
S	3.4	0.23	3.6	0.17	-1.20	3.4	0.17	3.4	0.14	0.06
29	358	0.21	364	0.25	-0.31	347	0.14	325	0.22	1.61
S	2.8	0.17	2.6	0.23	1.24	2.8	0.13	2.3	0.19	5.75X
51	311	0.22	328	0.16	-1.25	314	0.18	340	0.20	-1.84
S	2.4	0.18	2.4	0.12	0.56	2.5	0.13	2.4	0.15	1.67
35	286	0.21	310	0.17	-1.91	298	0.21	333	0.17	-2.67*
S	2.2	0.21	2.2	0.15	-0.09	2.4	0.20	2.3	0.15	0.50
30	256	0.23	338	0.16	-6.49*	289	0.29	238	0.23	3.29*
S	2.0	0.20	2.5	0.15	-5.37X	2.3	0.23	1.7	0.17	6.66X
24	218	0.23	221	0.23	-0.23	211	0.27	262	0.23	-3.98*
S	1.7	0.20	1.6	0.21	1.29	1.7	0.26	1.8	0.21	-1.76
37	243	0.18	199	0.21	4.53*	185	0.27	220	0.21	-3.27*
S	1.9	0.15	1.4	0.21	6.88X	1.5	0.25	1.5	0.18	-0.90
26	255	0.24	152	0.31#	8.41*	170	0.35#	229	0.24	-4.70*
S	2.0	0.20	1.1	0.28	10.94X	1.3	0.25	1.6	0.21	-3.65X
38	195	0.24	194	0.25	0.14	131	0.25	171	0.24	-4.75*
S	1.5	0.21	1.4	0.23	1.68	1.1	0.25	1.2	0.24	-2.34X
41	147	0.27	176	0.21	-3.31*	180	0.27	199	0.24	-1.68
S	1.1	0.25	1.3	0.18	-2.07X	1.4	0.24	1.4	0.21	0.68
40	137	0.25	155	0.26	-2.19*	178	0.29	174	0.25	0.43
S	1.1	0.23	1.1	0.26	-1.02	1.4	0.25	1.2	0.20	2.98X
50	137	0.27	152	0.26	-1.80	153	0.24	169	0.25	-1.92
S	1.1	0.25	1.1	0.24	-0.54	1.2	0.21	1.2	0.19	0.59
46	145	0.33#	117	0.25	3.08*	133	0.25	151	0.21	-2.49*
S	1.1	0.33%	0.9	0.23	4.18X	1.1	0.25	1.1	0.18	0.29
14	144	0.29	129	0.39#	1.44	129	0.39#	148	0.36#	-1.62
S	1.1	0.27	0.9	0.39%	2.46X	1.0	0.39%	1.0	0.33%	0.04
27	152	0.38#	112	0.23	3.96*	97	0.28	114	0.29	-2.54*
S	1.2	0.33%	0.8	0.21	5.43X	0.8	0.27	0.8	0.25	-0.39
52	106	0.22	102	0.33#	0.73	93	0.18	103	0.28	-1.89
S	0.8	0.20	0.7	0.32%	2.10X	0.7	0.20	0.7	0.25	0.85
33	47	0.55#	68	0.39#	-3.75*	76	0.32#	86	0.44#	-1.34
S	0.4	0.51%	0.5	0.36%	-3.30X	0.6	0.31%	0.6	0.37%	0.49
48	31	0.56#	38	0.24	-2.33*	26	0.28	24	0.39#	1.40
S	0.2	0.60%	0.3	0.23	-1.28	0.2	0.32%	0.2	0.39%	3.10X
T	12830	0.23	13801	0.22		12584	0.25	14303	0.23	
ST	100	0.20	100	0.21		100	0.22	100	0.19	

* The means are statistically different between the winter and spring of the year (X for shares)

Coefficient of variation (COV) at least 0.3 (% for shares)

T Total mean estimated local bus pass swipes for the garage and median COV (ST for shares)

four seasons in the table. All routes with high coefficients of variation are small routes with no more than 1.2% of the total estimated local bus pass swipes of the garage. However, there are large routes as well as small routes with mean estimated local bus pass swipe shares considered statistically different between the winter and the spring of 1994 (42%) and 1996 (38%). Therefore, we can conclude that this method produces relatively stable data (as shown by the shares) within a season on daily local bus pass swipes for large routes. However, since the method is incapable of producing stable data across seasons, even for the large routes, we have to conclude that Method 1 is incapable of satisfactorily estimating daily local bus pass swipes by route.

Table 4.17 shows the Method 1 results (including the shares) for estimating daily combo pass swipes by route.

Table 4.17 shows that five routes have high coefficients of variation (shares) in each season in 1994. There is some improvement in 1996 since only four routes in the winter and three in the spring have high coefficients of variation. Most of these routes are small or medium-sized routes. The improvement in 1996 is also consistent with the median coefficients of variation for all routes. In 1994, the medians are 0.22 for the winter and 0.21 for the spring, whereas the corresponding medians in 1996 are 0.21 and 0.19, respectively. In addition, many routes (all sizes) have mean estimated combo pass swipe shares statistically different between the winter and the spring (58% in 1994, and 46% in 1996). Therefore, there is improvement in the stability of data within a season between 1994 and 1996. However, like the local bus pass swipe shares analysis, there are large routes as well as small routes with mean estimated combo pass swipe shares considered statistically different between the winter and the spring of 1994 and 1996. Therefore, we can conclude that this method produces relatively stable data within a season on daily combo pass swipe shares for large routes when driver compliance is reasonably good.

Table 4.17: Daily Combo Pass Swipes: Method 1 Results

Rt	Winter Mean	COV	1994 Spring Mean	COV	Z-Val	Winter Mean	COV	1996 Spring Mean	COV	Z-Val
39	1368	0.12	1421	0.12	-1.46	1266	0.11	1418	0.09	-4.97*
S	15.7	0.09	15.9	0.08	-0.72	15.8	0.09	15.9	0.07	-0.29
32	972	0.16	1114	0.15	-3.98*	1039	0.17	1125	0.11	-2.50*
S	11.2	0.13	12.5	0.12	-3.95X	12.9	0.09	12.6	0.09	1.04
34	947	0.13	906	0.10	1.66	903	0.15	988	0.11	-3.10*
S	10.9	0.10	10.2	0.08	3.41X	11.2	0.11	11.1	0.09	0.72
28	840	0.15	971	0.16	-4.17*	787	0.16	931	0.15	-4.79*
S	9.6	0.11	10.8	0.12	-4.54X	9.8	0.08	10.4	0.09	-3.33X
31	857	0.14	916	0.15	-2.01*	779	0.25	910	0.16	-3.43*
S	9.8	0.11	10.2	0.11	-1.66	9.6	0.15	10.2	0.10	-1.99X
36	515	0.24	604	0.19	-3.31*	485	0.15	574	0.18	-4.42*
S	5.9	0.24	6.8	0.16	-3.01X	6.1	0.12	6.4	0.14	-1.89
51	390	0.21	393	0.17	-0.17	342	0.15	347	0.21	-0.34
S	4.5	0.18	4.4	0.13	0.52	4.3	0.14	3.9	0.16	3.03X
37	354	0.16	260	0.20	7.64*	230	0.26	278	0.21	-3.64*
S	4.1	0.15	2.9	0.19	8.97X	2.8	0.22	3.1	0.19	-2.00X
35	279	0.23	297	0.13	-1.53	287	0.18	298	0.17	-0.96
S	3.2	0.23	3.3	0.13	-0.95	3.6	0.19	3.3	0.15	1.93
29	281	0.19	282	0.23	-0.08	255	0.17	236	0.21	1.92
S	3.2	0.17	3.2	0.22	0.43	3.2	0.17	2.6	0.18	5.00X
26	333	0.24	172	0.37#	10.05*	218	0.38#	293	0.25	-4.24*
S	3.8	0.22	1.9	0.35%	11.18X	2.7	0.30%	3.3	0.22	-3.55X
30	236	0.19	284	0.22	-3.91*	236	0.30#	196	0.27	2.86*
S	2.7	0.18	3.2	0.16	-4.00X	2.9	0.25	2.2	0.22	5.32X
42	212	0.23	223	0.21	-1.06	180	0.19	201	0.30#	-1.96*
S	2.4	0.19	2.5	0.20	-0.80	2.2	0.17	2.2	0.27	0.04
50	187	0.27	148	0.26	3.93*	172	0.24	182	0.23	-1.07
S	2.1	0.24	1.7	0.25	4.61X	2.1	0.20	2.0	0.17	1.29
38	184	0.26	157	0.24	2.87*	128	0.27	160	0.22	-4.23*
S	2.1	0.22	1.8	0.24	3.57X	1.6	0.25	1.8	0.22	-2.30X
40	143	0.22	175	0.22	-4.01*	178	0.26	165	0.24	1.35
S	1.6	0.21	2.0	0.23	-3.60X	2.2	0.22	1.8	0.19	3.86X
52	163	0.21	170	0.36#	-0.62	131	0.13	156	0.28	-3.45*
S	1.9	0.19	1.9	0.32%	-0.15	1.6	0.15	1.7	0.23	-1.17
24	157	0.22	159	0.21	-0.23	135	0.23	176	0.21	-5.50*
S	1.8	0.22	1.8	0.21	0.25	1.7	0.24	2.0	0.19	-3.26X
27	134	0.31#	110	0.23	3.16*	105	0.27	119	0.26	-2.09*
S	1.5	0.26	1.2	0.21	3.93X	1.3	0.26	1.3	0.23	-0.16
41	47	0.34#	58	0.27	-2.91*	68	0.30#	59	0.27	2.22*
S	0.5	0.32%	0.6	0.24	-2.70X	0.8	0.28	0.7	0.26	4.00X
14	45	0.31#	42	0.56#	0.56	29	0.42#	36	0.41#	-2.08*
S	0.5	0.31%	0.5	0.57%	0.68	0.4	0.43%	0.4	0.40%	-0.85
46	40	0.66#	28	0.29	2.63*	38	0.27	40	0.25	-0.71
S	0.5	0.67%	0.3	0.27	2.82X	0.5	0.26	0.4	0.23	1.28
33	23	0.70#	39	0.43#	-4.37*	44	0.41#	50	0.40#	-1.30
S	0.3	0.67%	0.4	0.44%	-4.30X	0.6	0.43%	0.6	0.37%	0.10
48	14	1.24#	11	0.47#	1.18	10	0.34#	10	0.46#	-0.15
S	0.2	1.27%	0.1	0.44%	1.30	0.1	0.38%	0.1	0.48%	1.02
T	8721	0.23	8940	0.22		8045	0.25	8948	0.23	
ST	100	0.22	100	0.21		100	0.21	100	0.19	

* The means are statistically different between the winter and spring of the year (X for shares)

Coefficient of variation (COV) at least 0.3 (% for shares)

T Total mean estimated combo pass swipes for the garage and median COV (ST for shares)

However, since the method is incapable of producing stable data (shares) across seasons, even for the large routes, we have to conclude that Method 1 is incapable of satisfactorily estimating daily combo pass swipes by route.

There is one interesting phenomenon in this analysis: most high coefficients of variation and z-values are paired in the 1994 analysis of Table 4.17. This means that if a route has a high coefficient of variation for the estimated combo bus pass swipes, it is likely to have a high coefficient of variation for the estimated combo bus pass swipe share for the same season in 1994. This statement holds for the z-values in a similar manner. This is actually not a coincidence. From Table 4.15, we can see that the control totals of the 40-day sample for the combo pass analysis are not statistically different between the winter and the spring of 1994, which suggests that both the combo pass swipe analysis and the combo pass swipe share analysis should yield similar results. This is not true for other 40-day samples in Table 4.15 because the control totals are statistically different between the winter and the spring of a year. This shows the validity of analyzing estimated pass swipe shares instead of estimated pass swipes when the control totals are statistically different between the winter and the spring of a year.

4.3.2 Method 2: Allocation based on Missing Trip Records

Table 4.8 showed the average number of missing trip records per day for all four periods and the number of scheduled trips per day for each route. Table 4.18 shows the estimated local bus pass swipes and coefficients of variation for each route for both winter and spring for 1994 and 1996 by Method 2. As with the Method 1 analysis above, the estimated local bus pass swipe shares and the corresponding coefficients of variation are also shown. It also shows whether the means are statistically different between the winter and spring of each year (indicated by an * for mean estimated local bus pass swipes and an

Table 4.18: Daily Local Bus Pass Swipes: Method 2 Results

Rt	1994 Winter Mean	COV	1994 Spring Mean	COV	Z-Val	1996 Winter Mean	COV	1996 Spring Mean	COV	Z-Val
39	3620	0.11	3843	0.08	-2.85*	3374	0.10	3807	0.07	-6.33*
S	28.2	0.06	28.0	0.09	0.62	27.0	0.07	26.7	0.06	0.55
28	2003	0.13	2431	0.10	-7.56*	1975	0.16	2505	0.11	-7.79*
S	15.6	0.07	17.7	0.11	-5.99X	15.6	0.06	17.5	0.06	-8.70X
32	943	0.16	1041	0.13	-3.03*	1007	0.16	1094	0.10	-2.85*
S	7.4	0.15	7.5	0.10	-0.72	8.0	0.09	7.7	0.09	1.98X
31	840	0.14	890	0.15	-1.80	766	0.21	950	0.15	-5.49*
S	6.5	0.10	6.4	0.11	0.68	6.1	0.12	6.6	0.10	-3.90X
34	800	0.13	857	0.10	-2.65*	907	0.15	952	0.11	-1.62
S	6.2	0.08	6.2	0.08	0.21	7.2	0.13	6.7	0.10	3.23X
42	718	0.18	814	0.19	-3.08*	778	0.21	835	0.22	-1.46
S	5.6	0.12	5.9	0.16	-1.67	6.2	0.14	5.8	0.18	1.64
36	431	0.26	484	0.16	-2.48*	421	0.18	471	0.16	-2.93*
S	3.3	0.22	3.5	0.16	-1.15	3.3	0.13	3.3	0.12	0.63
26	426	0.21	317	0.73#	2.75*	317	0.41#	557	0.52#	-4.81*
S	3.3	0.16	2.2	0.72%	4.09X	2.5	0.31%	3.8	0.39%	-5.09X
29	354	0.20	357	0.25	-0.13	360	0.14	338	0.18	1.72
S	2.8	0.16	2.6	0.25	1.42	2.9	0.09	2.4	0.14	7.76X
30	306	0.21	334	0.16	-2.15*	357	0.27	270	0.20	5.07*
S	2.4	0.19	2.4	0.17	-0.53	2.8	0.18	1.9	0.14	10.29X
51	287	0.20	320	0.16	-2.69*	309	0.17	345	0.16	-3.07*
S	2.2	0.16	2.3	0.12	-1.13	2.5	0.10	2.4	0.11	0.82
35	284	0.19	298	0.17	-1.23	300	0.19	334	0.13	-3.06*
S	2.2	0.17	2.2	0.16	0.66	2.4	0.16	2.3	0.11	0.61
24	217	0.21	204	0.19	1.38	213	0.22	259	0.18	-4.47*
S	1.7	0.17	1.5	0.19	3.18X	1.7	0.18	1.8	0.15	-1.92
37	225	0.17	225	0.26	0.03	204	0.21	219	0.17	-1.62
S	1.8	0.14	1.6	0.24	1.81	1.6	0.20	1.5	0.14	1.50
14	243	0.33#	219	0.60#	0.98	139	0.34#	161	0.26	-2.17*
S	1.9	0.29	1.6	0.59%	1.96X	1.1	0.31%	1.1	0.22	-0.21
41	180	0.20	168	0.18	1.52	260	0.21	212	0.18	4.58*
S	1.4	0.21	1.2	0.18	3.23X	2.1	0.18	1.5	0.14	8.71X
38	173	0.22	207	0.31#	-2.92*	149	0.22	173	0.22	-3.04*
S	1.4	0.20	1.5	0.27	-1.83	1.2	0.20	1.2	0.21	-0.39
40	143	0.23	163	0.28	-2.26*	168	0.24	171	0.20	-0.38
S	1.1	0.21	1.2	0.25	-1.06	1.3	0.19	1.2	0.16	2.70X
50	132	0.26	150	0.26	-2.17*	151	0.22	173	0.19	-2.98*
S	1.0	0.24	1.1	0.24	-0.92	1.2	0.17	1.2	0.14	-0.23
46	138	0.33#	129	0.31#	0.94	125	0.21	145	0.19	-3.34*
S	1.1	0.32%	0.9	0.28	2.21X	1.0	0.21	1.0	0.15	-0.21
27	167	0.32#	119	0.25	4.98*	93	0.21	111	0.24	-3.57*
S	1.3	0.26	0.9	0.23	6.95X	0.7	0.20	0.8	0.19	-1.01
52	100	0.21	111	0.29	-1.73	93	0.17	99	0.24	-1.42
S	0.8	0.19	0.8	0.28	-0.48	0.7	0.18	0.7	0.22	1.62
33	65	0.50#	76	0.55#	-1.34	74	0.28	93	0.33#	-3.24*
S	0.5	0.48%	0.5	0.50%	-0.67	0.6	0.26	0.6	0.26	-1.46
48	38	0.53#	44	0.32#	-1.66	44	0.34#	28	0.41#	5.30*
S	0.3	0.56%	0.3	0.27	-0.71	0.3	0.28	0.2	0.39%	7.84X
T	12833	0.21	13801	0.22		12584	0.21	14302	0.19	
ST	100	0.18	100	0.21		100	0.18	100	0.14	

* The means are statistically different between the winter and spring of the year (X for shares)

Coefficient of variation (COV) at least 0.3 (% for shares)

T Total mean estimated local bus pass swipes for the garage and median COV (ST for shares)

X for mean estimated local bus pass swipe shares) and the z-values. In addition, any coefficient of variation of at least 0.3 is indicated by a # sign (% for shares). In this table routes are ordered by declining local bus pass swipes for ease of interpreting the impacts of route size.

The results are slightly better than those of Method 1, but the trends are quite similar. Table 4.18 shows that three routes in each season in 1994 and two in each season in 1996 have high coefficients of variation. Therefore, there is some improvement in the stability of data within a season between 1994 and 1996. Again, the median coefficients of variation in Table 4.18 confirm this. Other than the loop route 26, all routes are small or medium-sized routes with estimated local bus pass swipe shares less than 2%. However, like the results of Method 1, routes having statistically different estimated local bus pass swipe shares between the winter and the spring of the year (29% of the routes in 1994, and 42% of the routes in 1996) are not limited to the small ones. For example, route 28 has statistically different estimated local bus pass swipe shares in both 1994 and 1996. Therefore, we can conclude that this method produces relatively stable data within a season on daily local bus pass swipe shares for large routes. However, since the method is incapable of producing stable data across seasons, even for the large routes, we have to conclude that Method 2 is incapable of satisfactorily estimating daily local bus pass swipes by route.

Table 4.19 shows the Method 2 results on estimating daily combo pass swipes by route.

Table 4.19 shows that five routes in winter 1994, four in spring 1994, three in winter 1996, and two in spring 1996 have high coefficients of variation. This shows that there is consistent improvement between 1994 and 1996, and this is confirmed by the median coefficients of variation in the table. Again, other than the loop route 26, all of these routes

Table 4.19: Daily Combo Pass Swipes: Method 2 Results

Rt	Winter Mean	COV	1994 Spring Mean	COV	Z-Val	Winter Mean	COV	1996 Spring Mean	COV	Z-Val
39	1265	0.11	1292	0.10	-0.90	1163	0.10	1323	0.08	-6.28*
S	14.5	0.07	14.5	0.08	0.19	14.5	0.09	14.8	0.07	-1.14
32	987	0.15	1114	0.12	-3.93*	1003	0.16	1077	0.09	-2.52*
S	11.3	0.12	12.5	0.10	-3.99X	12.4	0.07	12.1	0.08	1.70
31	881	0.13	951	0.14	-2.53*	776	0.19	910	0.13	-4.57*
S	10.1	0.09	10.6	0.09	-2.40X	9.6	0.10	10.2	0.08	-2.93X
34	905	0.12	869	0.09	1.67	850	0.14	919	0.11	-2.90*
S	10.4	0.09	9.7	0.07	3.71X	10.6	0.10	10.3	0.09	1.32
28	805	0.14	919	0.14	-4.25*	737	0.15	895	0.13	-6.29*
S	9.2	0.10	10.3	0.10	-4.66X	9.1	0.07	10.0	0.08	-5.36X
36	497	0.23	572	0.17	-3.12*	469	0.14	548	0.16	-4.62*
S	5.7	0.22	6.4	0.14	-2.78X	5.9	0.10	6.1	0.12	-1.79
26	530	0.18	364	0.39#	6.16*	392	0.40#	586	0.35#	-4.74*
S	6.1	0.15	4.0	0.33%	7.91X	4.8	0.31%	6.4	0.26	-4.66X
51	357	0.19	378	0.17	-1.43	331	0.13	343	0.16	-1.12
S	4.1	0.16	4.2	0.12	-0.98	4.1	0.11	3.8	0.11	2.96X
37	324	0.16	292	0.20	2.59*	248	0.21	271	0.17	-2.17*
S	3.7	0.14	3.3	0.18	3.55X	3.1	0.17	3.0	0.15	0.34
35	272	0.21	283	0.13	-0.99	283	0.16	292	0.14	-0.98
S	3.1	0.20	3.2	0.13	-0.40	3.5	0.15	3.3	0.12	2.43X
30	277	0.21	272	0.20	0.40	285	0.27	211	0.22	5.17*
S	3.2	0.21	3.0	0.15	1.17	3.5	0.20	2.4	0.17	9.12X
29	274	0.19	276	0.19	-0.11	259	0.16	239	0.17	2.21*
S	3.1	0.16	3.1	0.17	0.52	3.2	0.15	2.7	0.13	6.09X
42	195	0.20	220	0.17	-2.86*	193	0.18	198	0.25	-0.53
S	2.2	0.16	2.5	0.17	-2.74X	2.4	0.14	2.2	0.22	2.13X
50	178	0.25	143	0.25	3.86*	167	0.21	181	0.18	-1.84
S	2.0	0.23	1.6	0.24	4.54X	2.1	0.17	2.0	0.13	0.79
40	146	0.19	184	0.20	-5.08*	165	0.21	161	0.20	0.64
S	1.7	0.18	2.1	0.18	-4.92X	2.1	0.17	1.8	0.15	3.65X
52	152	0.20	183	0.30#	-3.14*	129	0.14	149	0.25	-3.06*
S	1.7	0.18	2.0	0.27	-2.96X	1.6	0.14	1.7	0.21	-0.59
38	162	0.25	159	0.21	0.45	142	0.22	159	0.19	-2.44*
S	1.9	0.22	1.8	0.22	0.89	1.8	0.19	1.8	0.19	-0.13
24	153	0.18	142	0.19	1.71	134	0.18	171	0.17	-6.34*
S	1.8	0.18	1.6	0.19	2.15X	1.7	0.17	1.9	0.14	-3.90X
27	146	0.26	121	0.26	3.24*	99	0.22	114	0.22	-2.83*
S	1.7	0.21	1.3	0.22	4.33X	1.2	0.21	1.3	0.18	-0.62
41	57	0.31#	54	0.25	0.79	96	0.24	61	0.20	8.44*
S	0.6	0.31%	0.6	0.22	1.28	1.2	0.21	0.7	0.19	11.29X
14	73	0.37#	68	0.54#	0.74	31	0.34#	37	0.32#	-2.52*
S	0.8	0.35%	0.8	0.55%	0.89	0.4	0.33%	0.4	0.30%	-1.06
33	31	0.66#	43	0.35#	-3.20*	42	0.38#	52	0.30#	-2.91*
S	0.4	0.64%	0.5	0.35%	-3.08X	0.5	0.40%	0.6	0.26	-1.35
46	37	0.69#	31	0.28	1.57	36	0.24	38	0.23	-1.30
S	0.4	0.69%	0.3	0.24	1.83	0.4	0.23	0.4	0.21	0.91
48	17	1.18#	13	0.44#	1.29	16	0.30#	11	0.44#	4.36*
S	0.2	1.21%	0.1	0.41%	1.40	0.2	0.27	0.1	0.45%	5.85X
T	8721	0.20	8943	0.20		8046	0.20	8946	0.19	
ST	100	0.18	100	0.18		100	0.17	100	0.15	

* The means are statistically different between the winter and spring of the year (X for shares)

Coefficient of variation (COV) at least 0.3 (% for shares)

T Total mean estimated combo pass swipes for the garage and median COV (ST for shares)

are small routes with less than 1% of estimated combo pass swipe shares. However, as in previous analyses, routes having statistically different mean estimated combo pass swipe shares between the winter and spring of the year (58% in 1994 and 50% in 1996) include both large and small ones. Therefore, there is overall improvement in performance for this method between 1994 and 1996, but there is no significant improvement over Method 1. We can therefore conclude that Method 2 is incapable of satisfactorily estimating daily combo pass swipes by route. Finally, as in Table 4.17, we can see that the coefficients of variation and z-values are also paired in the 1994 analysis of Table 4.19 because the control totals are not statistically different between the winter and the spring of 1994 as shown in Table 4.15.

4.3.3 Method 3: Screen Bad Trip Records then Apply Method 2

The analysis of Method 3 is different from that of Methods 1 and 2 in three ways. First, only 5 days are analyzed in each season, so the control totals are different from those of Methods 1 and 2 (please refer to Figure 4.12). Second, since the sample size is small, t-values instead of z-values are used. Finally, route 34E is analyzed separately from route 34.

4.3.3.1 Trip Screening Results

Table 4.20 shows a summary of the eliminations performed. The first half of the figure shows the eliminations performed on 1994 data whereas the second half shows the same information for 1996 data. The trends are very similar to those in Table 4.10 for the fare revenue analysis because they are based on the same seasons, namely the winter and the spring of 1994 and 1996; however, the two tables are not identical because the fare

Table 4.20: Daily Pass Swipes: A Summary of Eliminations Performed in Method 3

1994 Data:

Route	SchTrip	Trips	Defects	Short	Long	Remaining	(% of SchTrip)
39	4010	2647	3	277	905	1462	(36%)
32	2410	1124	2	105	414	603	(25%)
28	2350	1323	3	146	530	644	(27%)
31	2060	878	0	78	309	491	(24%)
34	1380	737	1	54	275	407	(29%)
36	1360	747	0	59	245	443	(33%)
42	1190	705	0	82	176	447	(38%)
29	890	461	1	45	193	222	(25%)
30	880	380	0	29	143	208	(24%)
41	870	422	0	73	87	262	(30%)
35	840	513	0	44	166	303	(36%)
34E	820	536	1	76	212	247	(30%)
37	800	425	0	23	140	262	(33%)
24	710	439	1	38	114	286	(40%)
38	660	425	0	71	118	236	(36%)
51	640	395	3	5	131	256	(40%)
26	580	84	0	13	64	7	(1%)
50	560	305	0	12	95	198	(35%)
46	510	308	0	9	58	241	(47%)
52	470	255	0	27	112	116	(25%)
27	440	223	0	6	71	146	(33%)
40	440	158	0	10	65	83	(19%)
33	280	106	0	26	42	38	(14%)
14	270	56	0	2	34	20	(7%)
48	150	55	0	4	36	15	(10%)
Total:	25570	13707	15	1314	4735	7643	(30%)
% of SchTrip:		54%	0%	5%	19%	30%	

1996 Data:

Route	SchTrip	Trips	Defects	Short	Long	Remaining	(% of SchTrip)
39	4010	3888	12	311	909	2656	(66%)
32	2410	2167	8	114	424	1621	(67%)
28	2350	2181	0	219	591	1371	(58%)
31	2060	1779	0	121	326	1332	(65%)
34	1380	1115	1	59	317	738	(53%)
36	1360	1199	1	76	337	785	(58%)
42	1190	973	0	119	201	653	(55%)
29	890	765	0	45	239	481	(54%)
30	880	555	1	40	173	341	(39%)
41	870	606	0	19	118	469	(54%)
35	840	740	0	49	165	526	(63%)
34E	820	1309	0	203	443	663	(81%)
37	800	685	0	32	170	483	(60%)
24	710	659	0	45	194	420	(59%)
38	660	525	0	17	123	385	(58%)
51	640	534	0	13	96	425	(66%)
26	580	181	0	29	97	55	(9%)
50	560	492	0	24	103	365	(65%)
46	510	477	0	4	51	422	(83%)
52	470	398	0	61	97	240	(51%)
27	440	404	0	8	45	351	(80%)
40	440	406	0	10	86	310	(70%)
33	280	228	0	33	55	140	(50%)
14	270	204	0	9	33	162	(60%)
48	150	88	9	8	33	38	(25%)
Total:	25570	22558	32	1668	5426	15432	(60%)
% of SchTrip:		88%	0%	7%	21%	60%	

revenue and pass swipe analyses are based on different days. Also, the definition of a defective bus is different for the two analyses. For the fare revenue analysis, a defective bus is defined as a bus having a farebox failing to count cash fare whereas a defective farebox in the pass swipe analysis is defined as a bus having a defective pass reader. Again, please refer to Table 3.2 for the median inter-record times for each routes in Bartlett Garage calculated from both the 1994 and 1996 data. These are used for the first and second rounds of screening.

From Table 4.20, a number of observations can be made. First, in 1994 the overall number of recorded trips is only slightly more than half (54%) of the number of scheduled trips. There are only 13707 recorded trips out of 25570 scheduled trips in Bartlett Garage in the 10 days analyzed. This means that many operators were either not registering at all or only registering one trip out of many trips. This is confirmed by the number of trips eliminated because they are spaced too far apart from the next trip. A total of 4735 out of 13707 recorded trips (35%) are eliminated due to this reason. However, there is dramatic improvement in 1996. A total of 22558 (88%) out of 25570 trips are recorded, and only 5426 out of 22558 recorded trips (24%) are eliminated because they are spaced too far apart from the next trip. This even outperforms the five days used in the fare revenue analysis.

Second, multiple registration is not as serious a problem as only 1314 out of 13707 recorded trips (10%) are eliminated in 1994 for this reason. There is also improvement in 1996 regarding multiple registration. Only 1668 out of 22558 trips (7%) have their revenue accumulated with the next trip because they are spaced too close to the next trip.

Third, only 15 out of 13707 recorded trips are eliminated due to defective fareboxes. This problem is slightly worse with 1996 data than with 1994 data since 32 out of 22558 trips are eliminated. However, this is still much better than the situation in the fare revenue analysis. Based on our analysis, since this problem is not related to driver compliance, we can conclude that the pass readers are generally more reliable than the coin collectors on electronic fareboxes.

Fourth, we recall from the discussion above that the routes 14, 26, 33, and 48 have been quite problematic, especially in 1994, when they often had high coefficients of variation. This can be explained by Table 4.20. All of these problem routes have the number of remaining trips less than 15% of the number of scheduled trips in 1994, while the average for the garage is 30%. This shows that there is a dearth of good data for these problem routes. In fact, many of these trips have very few recorded trips to start with. Route 26 is the most serious one. It has only 84 recorded trips out of a total of 580 scheduled trips (15%), and after the eliminations, it has only 7 trips left; only 1% of the scheduled trips. Similarly, for routes 14, 33, and 48, the percentage of scheduled trips recorded (before the eliminations) are only 21%, 38%, and 37%, respectively, which are significantly less than the average of 54%. This outcome is very similar to that of the fare revenue analysis. As described in the fare revenue analysis, one reason for the poor performance of routes 26 and 48 is that they are loop routes. Some operators may not know how to register with the farebox properly on loop routes. Fortunately, all of these routes have improved in 1996, with the improvement for routes 14 and 33 especially significant.

Finally, in 1994 the number of remaining trips is 7643, which is only about 56% of the number of recorded trips and 30% of the number of scheduled trips. This means that there are not many trips left after the eliminations, especially for the small routes. Fortunately, this situation has also improved dramatically in 1996. There are a total of 15432 trips left after all eliminations in 1996. This is 68% of the number of trips recorded and 60% of the

Table 4.21: Daily Pass Swipes: Trip Record Screening Results

		1994				1996			
Number of Scheduled Trips > 100									
Route	SCH_T	W_TPS	0_DYS	S_TPS	0_DYS	W_TPS	0_DYS	S_TPS	0_DYS
391	401	113	0	179	0	255	0	276	0
32	241	42	0	78	0	161	0	164	0
28c	235	51	0	76	0	142	0	132	0
31	206	40	0	58	0	135	0	131	0
34	138	34	0	47	0	62	0	86	0
36c	136	39	0	50	0	80	0	77	0
42C	119	47	0	42	0	59	0	72	0
Number of Scheduled Trips between 50 and 100 inclusive									
Route	SCH_T	W_TPS	0_DYS	S_TPS	0_DYS	W_TPS	0_DYS	S_TPS	0_DYS
29	89	20	0	24	0	45	0	51	0
30LC	88	12	0	30	0	29	0	39	0
41Cc	87	16	0	37	0	42	0	51	0
351	84	29	0	32	0	53	0	52	0
34El	82	20	0	29	0	57	0	76	0
37c	80	33	0	20	0	45	0	52	0
24	71	22	0	35	0	39	0	45	0
38	66	25	0	22	0	37	0	40	0
51	64	30	0	21	0	40	0	45	0
26LlCc	58	0	5	1	2	8	1	3	0
50c	56	20	0	20	0	35	0	38	0
46L	51	29	0	19	0	40	0	44	0
Number of scheduled trips < 50									
Route	SCH_T	W_TPS	0_DYS	S_TPS	0_DYS	W_TPS	0_DYS	S_TPS	0_DYS
52LCc	47	17	0	6	0	22	0	26	0
27C	44	13	0	16	0	34	0	36	0
40LC	44	9	0	8	0	32	0	30	0
33LCc	28	3	0	5	1	16	0	12	0
14LlCc	27	1	4	3	0	16	0	17	0
48LlCc	15	0	5	3	0	3	1	4	0

L Coefficient of variation (COV) at least 0.3 (or unavailable) in one or more of the four seasons in the local bus pass swipe analysis

l The means are statistically different between the winter and spring of the year (or T value unavailable) in one or more of the two years in the local bus pass swipe analysis

C Coefficient of variation (COV) at least 0.3 (or unavailable) in one or more of the four seasons in the combo pass swipe analysis

c The means are statistically different between the winter and spring of the year (or T value unavailable) in one or more of the two years in the combo pass swipe analysis number of scheduled trips, which is a huge improvement from 1994. In addition, except for route 26 (a problem route), all routes have more than 25% of the number of scheduled trips remaining after all eliminations. This is better than that of the fare revenue analysis. This is very useful for the small routes because now they have enough good data to provide a reasonable estimate of daily pass swipes. The five days chosen for the pass swipe analysis in 1996 are later in the year than those chosen for the fare revenue analysis, which further confirms that driver compliance continues to improve with time.

Therefore, these observations further confirm our findings in the fare revenue analysis that Method 3 should work well with all routes except the really small and looped ones such as routes 26 and 48.

Table 4.21 shows the number of trip records remaining after the screenings, summarized by the number of scheduled trips per day. Similar to Table 4.11 in the fare revenue analysis, Table 4.21 is divided into three sections: the large routes, the medium-sized routes, and the small routes.

From Table 4.21, we can see that, similar to the fare revenue analysis, all large and medium-sized routes except the loop route 26 have data for all days analyzed. For the small routes, 3 out of 6 routes (routes 14, 33, and 48) have no data for one or more days analyzed in 1994. However, in 1996, only route 48 has no data for one day in the winter. This shows that for pass swipe analysis, Method 3 eliminations generally work very well with large and medium-sized routes, and with good driver compliance, it can work very well with small routes too.

Table 4.22: Daily Local Bus Pass Swipes: Method 3 Results

Rt	1994 Winter Mean	COV	1994 Spring Mean	COV	T-Val	1996 Winter Mean	COV	1996 Spring Mean	COV	T-Val
39	3382	0.16	4263	0.07	-3.18*	4192	0.08	4058	0.04	0.84
S	26.6	0.06	29.0	0.02	-3.30X	28.4	0.03	28.2	0.02	0.38
32	965	0.22	1137	0.13	-1.49	1227	0.09	1232	0.02	-0.10
S	7.5	0.12	7.7	0.08	-0.34	8.3	0.03	8.6	0.03	-1.64
28	2376	0.12	2558	0.13	-0.92	2452	0.11	2595	0.03	-1.19
S	18.8	0.06	17.3	0.06	2.07	16.6	0.05	18.1	0.01	-3.78X
31	805	0.13	868	0.09	-1.06	934	0.09	898	0.05	0.83
S	6.4	0.13	5.9	0.09	1.08	6.3	0.04	6.2	0.03	0.50
34	457	0.18	516	0.09	-1.38	562	0.17	517	0.07	1.02
S	3.6	0.07	3.5	0.12	0.29	3.8	0.09	3.6	0.08	0.94
36	449	0.16	532	0.11	-2.04	452	0.13	455	0.07	-0.12
S	3.6	0.17	3.6	0.10	-0.19	3.1	0.05	3.2	0.05	-1.17
42	753	0.24	962	0.11	-2.25	933	0.14	831	0.04	1.64
S	5.9	0.13	6.5	0.07	-1.66	6.3	0.07	5.8	0.03	2.57X
29	460	0.16	507	0.19	-0.86	467	0.21	376	0.09	1.96
S	3.6	0.14	3.4	0.15	0.63	3.1	0.15	2.6	0.09	2.27
30	326	0.43#	359	0.14	-0.49	321	0.18	284	0.05	1.37
S	2.6	0.46%	2.4	0.12	0.28	2.2	0.11	2.0	0.05	1.55
41	209	0.20	181	0.18	1.17	244	0.19	218	0.10	1.15
S	1.6	0.12	1.2	0.12	3.82X	1.7	0.17	1.5	0.08	1.02
35	382	0.13	282	0.21	2.97*	347	0.08	373	0.04	-1.76
S	3.0	0.12	1.9	0.20	4.55X	2.4	0.08	2.6	0.04	-2.55X
34E	312	0.17	450	0.16	-3.50*	264	0.27	247	0.15	0.46
S	2.5	0.27	3.1	0.15	-1.53	1.8	0.20	1.7	0.15	0.25
37	216	0.06	214	0.20	0.15	255	0.08	230	0.10	1.85
S	1.7	0.07	1.4	0.17	2.24	1.7	0.05	1.6	0.09	1.76
24	211	0.21	226	0.17	-0.56	296	0.19	240	0.09	2.09
S	1.7	0.13	1.5	0.12	1.02	2.0	0.14	1.7	0.09	2.29
38	210	0.23	174	0.20	1.38	183	0.15	168	0.09	1.10
S	1.6	0.14	1.2	0.13	3.76X	1.2	0.11	1.2	0.09	0.90
51	318	0.12	352	0.19	-1.01	389	0.15	406	0.07	-0.60
S	2.5	0.07	2.4	0.17	0.60	2.6	0.15	2.8	0.05	-1.01
26	NA	NA#	278	0.89#	NA*	204	0.19	194	0.48#	0.24
S	NA	NA%	2.0	0.89%	NAX	1.4	0.14	1.3	0.46%	0.23
50	141	0.10	163	0.17	-1.62	185	0.18	174	0.11	0.64
S	1.1	0.07	1.1	0.10	0.19	1.2	0.13	1.2	0.11	0.37
46	183	0.34#	134	0.19	1.63	160	0.07	141	0.18	1.52
S	1.4	0.21	0.9	0.14	3.55X	1.1	0.07	1.0	0.15	1.45
52	128	0.15	100	0.25	1.97	117	0.44#	95	0.21	0.88
S	1.0	0.18	0.7	0.24	3.01X	0.8	0.37%	0.7	0.19	0.82
27	105	0.10	107	0.15	-0.22	116	0.12	111	0.08	0.70
S	0.8	0.19	0.7	0.08	1.56	0.8	0.09	0.8	0.07	0.37
40	174	0.25	164	0.46#	0.25	189	0.11	181	0.03	0.76
S	1.4	0.34%	1.1	0.39%	1.10	1.3	0.13	1.3	0.05	0.26
33	98	0.64#	104	0.09	-0.21	116	0.14	127	0.63#	-0.32
S	0.8	0.52%	0.7	0.12	0.16	0.8	0.20	0.9	0.65%	-0.37
14	108	NA#	170	0.42#	NA*	190	0.16	199	0.12	-0.55
S	0.8	NA%	1.2	0.38%	NAX	1.3	0.15	1.4	0.13	-0.85
48	NA	NA#	36	0.42#	NA*	31	0.77#	17	0.35#	1.32
S	NA	NA%	0.2	0.45%	NAX	0.2	0.69%	0.1	0.34%	1.41
T	12768	0.18	14837	0.17		14826	0.15	14367	0.09	
ST	100	0.14	100	0.12		100	0.11	100	0.08	

* The means are statistically different between the winter and spring of the year (X for shares)
# Coefficient of variation (COV) at least 0.3 (% for shares)
T Total mean estimated local bus pass swipes for the garage and median COV (ST for shares)

4.3.3.2 Analysis Results

Table 4.22 shows the results of Method 3 on local bus pass swipes. As in Table 4.12, the routes are classified into 3 categories. Again, the estimated local bus pass swipe shares are also presented. However, from Table 4.15, we can see that the control totals for 1996 are not statistically different; therefore, the direct analysis of estimated local bus pass swipes is valid. The estimated local bus pass swipe shares are provided for consistency.

For the large routes (refer to Table 4.21 for the classification of routes by size), this method performs extremely well. First of all, no coefficient of variation among routes with more than 100 scheduled trips each day is considered high. In fact, the highest coefficient of variation is only 0.24 overall, and only 0.07 in the spring of 1996. The estimated local bus swipe shares for large routes have an overall high of 0.17, and a high of 0.08 in the spring of 1996. Furthermore, in 1994, only one large route was labelled different. In 1996, only the estimated shares of two large routes are labelled different; however, this is because of extremely low coefficients of variation in both cases. We can see that this method performs extremely well on large routes, and that the data within each season are very stable.

Only three medium-sized routes have one or more coefficients of variation of at least 0.3. They are routes 26, 30, and 46. For the shares, only two routes have high coefficients of variation (routes 26 and 30). In fact, route 26 has one or more days with no trip record remaining in three out of four seasons, and its coefficients of variation are generally quite high except in the winter of 1996 (0.89 in spring 1994 and 0.48 in spring 1996). This means that route 26 also cannot be properly analyzed by this method. However, since only 3 out of 12 routes are labelled different (routes 26, 34E, and 35), and two more using the share analysis (routes 26, 35, 38, 41, and 46), we can say that this method performs reasonably well on medium-sized routes and that the data within each season are generally quite stable. Furthermore, with 1996 data, other than the problem route 26, and route 35, all the remaining 10 medium-sized routes have satisfactory results. Hence we can see that Method 3 works very well with medium-sized routes given good driver compliance.

A total of 3 out of 25 routes have NA labels with 1994 data. They are routes 14, 26, and 48, which also have NA labels in the daily fare revenue analysis above. This automatically means that they cannot be properly analyzed by this method in 1994. One similarity among them is that they are all relatively small routes, most of them with fewer than 50 scheduled trips each day. Although route 26 has 58 scheduled trips per day, it has very few trip records remaining after the screenings. However, no NA label appears in the 1996 analysis. This means that all routes, even the very small ones, have data remaining after all eliminations. This is quite an improvement over the 1994 data. In addition, we can see that two small routes (routes 27 and 52) have satisfactory results in 1994. That is, these two routes have both coefficients of variation less than 0.3 and have mean estimated local bus pass swipes that are statistically indifferent between the winter and the spring of the year. For the share analysis, only route 27 has satisfactory results. In 1996, three small routes (routes 14, 27, and 40) have satisfactory results in both the pass swipe and share analyses. Again, this is an improvement over 1994. Furthermore, while two out of six small routes (three out of six for the share analysis) are considered to have statistically different mean estimated local bus pass swipes between the winter and the spring of 1994, none of the small routes are considered to have statistically different means in 1996.

Therefore, we can conclude that the data within each season are not yet stable for small routes under Method 3. However, as mentioned in the fare revenue analysis, remember that we are being conservative in this analysis. That is, we would rather reject a good method than to accept a bad method. The very small sample size for Methods 3 and

4 implies that we can place very little confidence in the coefficient of variation figures, and it is possible that the coefficients of variation will change by increasing the sample size. We use 0.3 as the cutoff here simply as a screening tool, but the implications of the very small sample size should not be overlooked. With improved driver compliance, there is potential that the estimated local bus pass swipes be very stable within a season even for small routes. Together with the decrease in the percentage of small routes with means considered statistically different, this suggests that there is potential for Method 3 to be used to estimate daily local bus pass swipes by route if driver compliance continues to improve over time.

Overall, 36% of the routes in 1994, and 12% in 1996 have statistically different estimated local bus pass swipe shares. Although it is still lower than the 10% preset threshold, this shows that Method 3 is very promising in estimating local bus pass swipes by route, at least for the medium-sized (except route 26) and large routes.

The success of Method 3 with 1996 data is also due to the high quality of the 1996 data. First, there are generally fewer days with no data remaining in 1996 than in 1994, which in turn leads to fewer NA labels (in this case, no NA labels) in 1996. In the spring of 1996, there is no route with no data for any single day analyzed, which is especially important for the small routes. Second, the coefficients of variation in 1996 are generally lower than those in 1994. In 1996, only 5 coefficients of variation in both the pass swipe and share analyses are greater than 0.3, and the highest one is 0.77 (0.69 in the share analysis) for route 48 in the winter. On the other hand, for both the pass swipe and share analyses in 1994, other than the ones unavailable, 7 coefficients of variation are greater than 0.3, the highest being 0.89 for route 26 in the spring. Finally, none of the routes in 1996 have statistically different means (and only two have statistically different mean estimated local bus swipe shares) whereas 24% of the routes (36% of the routes for shares)

Table 4.23: Daily Combo Pass Swipes: Method 3 Results

Rt	1994 Winter Mean	COV	1994 Spring Mean	COV	T-Val	1996 Winter Mean	COV	1996 Spring Mean	COV	T-Val
39	1176	0.15	1385	0.08	-2.29	1440	0.11	1390	0.06	0.64
S	13.5	0.12	14.3	0.05	-1.05	15.5	0.06	15.4	0.04	0.30
32	1196	0.19	1200	0.10	-0.03	1263	0.05	1330	0.02	-2.15
S	13.6	0.11	12.4	0.08	1.44	13.7	0.08	14.7	0.05	-1.82
28	898	0.07	1043	0.09	-2.89*	895	0.06	924	0.10	-0.62
S	10.3	0.06	10.8	0.09	-0.97	9.7	0.07	10.2	0.10	-0.98
31	862	0.17	867	0.13	-0.07	915	0.12	826	0.07	1.59
S	9.8	0.09	9.0	0.11	1.48	9.9	0.08	9.1	0.06	1.62
34	779	0.16	731	0.15	0.66	664	0.12	639	0.09	0.59
S	8.9	0.13	7.5	0.12	2.07	7.2	0.03	7.1	0.10	0.22
36	439	0.14	556	0.14	-2.68*	531	0.12	520	0.07	0.34
S	5.0	0.13	5.7	0.08	-1.92	5.7	0.09	5.7	0.04	-0.04
42	184	0.33#	242	0.11	-1.99	228	0.11	210	0.05	1.52
S	2.1	0.27	2.5	0.11	-1.53	2.5	0.09	2.3	0.07	1.13
29	415	0.27	374	0.24	0.62	320	0.19	256	0.08	2.20
S	4.7	0.25	3.9	0.22	1.38	3.4	0.11	2.8	0.09	2.84X
30	233	0.58#	329	0.12	-1.52	286	0.25	257	0.19	0.74
S	2.6	0.53%	3.4	0.07	-1.26	3.0	0.17	2.8	0.15	0.72
41	75	0.32#	62	0.19	1.10	105	0.18	71	0.16	3.41*
S	0.8	0.23	0.6	0.21	1.92	1.1	0.19	0.8	0.13	3.29X
35	307	0.09	352	0.22	-1.21	315	0.16	283	0.06	1.38
S	3.5	0.15	3.6	0.16	-0.19	3.4	0.08	3.1	0.08	1.42
34E	338	0.17	367	0.13	-0.84	206	0.10	187	0.12	1.41
S	3.9	0.19	3.8	0.09	0.31	2.2	0.12	2.1	0.12	0.95
37	364	0.12	294	0.16	2.46*	308	0.22	300	0.21	0.19
S	4.2	0.16	3.0	0.15	3.19X	3.3	0.14	3.3	0.20	-0.03
24	185	0.12	178	0.06	0.65	154	0.13	155	0.14	-0.07
S	2.1	0.07	1.8	0.06	3.27X	1.7	0.07	1.7	0.11	-0.51
38	166	0.22	185	0.22	-0.78	165	0.18	180	0.17	-0.78
S	1.9	0.16	1.9	0.18	-0.10	1.8	0.12	2.0	0.15	-1.31
51	383	0.09	447	0.14	-1.99	416	0.13	422	0.04	-0.24
S	4.4	0.09	4.6	0.10	-0.79	4.5	0.11	4.7	0.03	-0.76
26	NA	NA#	472	1.11#	NA*	322	0.44#	274	0.57#	0.51
S	NA	NA%	5.0	1.11%	NAX	3.5	0.35%	3.0	0.54%	0.53
50	225	0.27	159	0.08	2.44*	204	0.07	205	0.09	-0.10
S	2.6	0.21	1.6	0.02	3.88X	2.2	0.10	2.3	0.10	-0.42
46	36	0.23	35	0.27	0.21	43	0.20	41	0.27	0.35
S	0.4	0.31%	0.4	0.28	0.78	0.5	0.16	0.5	0.27	0.18
52	202	0.10	168	0.14	2.46*	142	0.30#	145	0.27	-0.09
S	2.3	0.14	1.7	0.16	3.06X	1.5	0.24	1.6	0.23	-0.33
27	91	0.30	87	0.19	0.33	120	0.13	116	0.33#	0.21
S	1.0	0.28	0.9	0.22	0.95	1.3	0.13	1.3	0.32%	0.08
40	136	0.47#	222	0.26	-2.25	185	0.18	180	0.08	0.33
S	1.6	0.52%	2.3	0.30%	-1.53	2.0	0.10	2.0	0.09	-0.05
33	51	0.40#	88	0.21	-3.02*	63	0.13	87	1.14#	-0.55
S	0.6	0.35%	0.9	0.18	-2.77X	0.7	0.16	1.0	1.16%	-0.59
14	28	NA#	40	0.58#	NA*	38	0.50#	39	0.10	-0.13
S	0.3	NA%	0.4	0.58%	NAX	0.4	0.45%	0.4	0.12	-0.32
48	NA	NA#	7	1.38#	NA*	13	1.34#	5	0.89#	1.04
S	NA	NA%	0.1	1.38%	NAX	0.1	1.27%	0.1	0.89%	1.05
T	8769	0.22	9890	0.15		9341	0.13	9042	0.10	
ST	100	0.19	100	0.15		100	0.11	100	0.11	

* The means are statistically different between the winter and spring of the year (X for shares)
# Coefficient of variation (COV) at least 0.3 (% for shares)
T Total mean estimated combo pass swipes for the garage and median COV (ST for shares)

do so in 1994. This further shows that the method has the potential to perform very well if the data continue to improve.

Table 4.23 shows the results of Method 3 on daily combo pass swipes by route. Again, the routes are classified into 3 categories. From Table 4.15, we can see that the control totals for combo pass swipes are not statistically different between the winter and the spring of both 1994 and 1996, hence the estimated combo pass swipe analysis (without using shares) is completely valid. Although the shares are listed for consistency in Table 4.23, the discussion below focuses on the estimated combo pass swipes themselves.

We can see from Table 4.23 that with the large routes (refer to Table 4.21 for the classification of routes by size), this method performs very well. First of all, only one route (route 42) among routes with more than 100 scheduled trips each day has a coefficient of variation considered to be high. In fact, other than the high coefficient of variation of route 42 in the winter of 1994 (0.33), the highest coefficient of variation is only 0.19 overall, and only 0.10 in the spring of 1996. Furthermore, although two large routes are labelled statistically different in 1994, no large route is labelled different in 1996. We can see that, with good driver compliance, this method performs extremely well on large routes, and that the data within each season are very stable.

Three medium-sized routes have one or more coefficients of variation of at least 0.3. They are routes 26, 30, and 41. In fact, route 26 has one or more days with no trip record remaining in three out of four seasons, and its coefficient of variation is either high or unavailable. This means that route 26 also cannot be properly analyzed by this method. However, since only 4 out of 12 routes are labelled different (routes 26, 37, 41, and 50), 7 routes (excluding route 30 which has a high coefficient of variation) are still considered appropriate for this method. We can see that this method performs reasonably well on medium-sized routes and that the data within each season are generally quite stable.

Furthermore, with 1996 data, other than the problem route 26, and another route 41, all the remaining 10 medium-sized routes have satisfactory results. Hence we can see that Method 3 works very well with medium-sized routes given good driver compliance.

We can see from Table 4.23 that 3 of the 25 routes have NA labels with 1994 data. They are the familiar problem routes 14, 26, and 48. This automatically means that they cannot be properly analyzed by this method in 1994. However, no NA label appears in the 1996 analysis. This means that all routes, even the very small ones, have data remaining after all eliminations. This is quite an improvement over the 1994 data. Furthermore, while four out of six small routes are considered to have statistically different mean estimated combo pass swipes between the winter and the spring of 1994, none of the small routes are considered to have statistically different means in 1996. However, only one small route in each year (route 27 in 1994, route 40 in 1996) has satisfactory results. Route 52 in the winter of 1996 is a near-miss case. The coefficient of variation is 0.30, which is right at the threshold. However, as mentioned before, if the sample size is increased, it is possible that the coefficients of variation of some of these small routes will fall below 0.3. We can conclude that the data within each season are not yet stable for small routes under Method 3. However, with improved driver compliance, there is potential that the estimated combo pass swipes be very stable within a season even for small routes. Together with the decrease in the percentage of small routes with means considered statistically different, this suggests that there is potential for Method 3 to be used to estimate daily combo pass swipes by route if driver compliance continues to improve over time.

Overall, 36% of the routes in 1994 and 4% in 1996 have estimated combo pass swipes statistically different between the winter and the spring of the year. The results of 1996 are very encouraging. Besides the relatively small number of routes have high coefficients of variation, the fact that all of these routes are relatively small routes (with estimated daily combo pass swipes around 200 at most except the problem route 26) is also important. Furthermore, from Table 4.23, we can see that most large and medium-sized routes (except routes 26, 29 (for share analysis), and 41) have satisfactory results in 1996. This tells us that most medium-sized and large routes can be satisfactorily analyzed by Method 3.

As in all our previous analyses with fare revenue and local bus pass swipes, the 1996 sample is of much higher quality than its 1994 counterpart. First, there are generally fewer days with no data remaining in 1996 than in 1994, which in turn leads to fewer NA labels (in this case, no NA labels) in 1996. In the spring of 1996, there is no route with no data for any single day analyzed, which is especially important for the small routes. Second, the coefficients of variation in 1996 are generally lower than those in 1994. In 1996, only 8 coefficients of variation are greater than 0.3 compared with 11 coefficients of variation in 1994. This further shows that the method has the potential to perform very well if the data continue to improve.

4.3.4 A Comparison of the Three Methods

This section compares the three methods presented above for estimating daily pass swipes by route. Since most of the pass swipe analyses use estimated pass swipe shares instead of estimated pass swipes, the comparisons in this section will focus on estimated pass swipe shares.

Table 4.24 shows the comparison of estimated local bus pass swipe shares, coefficients of variation, percentage difference in estimated local bus pass swipe shares for the two seasons, and t/z-values produced by the three different methods based on the 1996 data. Again, the 1994 results are not compared because the 1996 results are much more relevant

Table 4.24: Local Bus Pass Swipes: A Comparison of the Three Methods (1996 data)

1996 Data
Local Bus Pass Swipe Shares

Number of Scheduled Trips > 100

	METHOD 1	METHOD 2	METHOD 3
Route: 39			
Winter:	28.9/0.07	27.0/0.07	28.4/0.03
Spring:	28.4/0.06	26.7/0.06	28.2/0.02
% Difference:	-2	-1	-1
T/Z-Value:	1.05	0.55	0.38
Route: 32			
Winter:	8.1/0.10	8.0/0.09	8.3/0.03
Spring:	7.9/0.10	7.7/0.09	8.6/0.03
% Difference:	-2	-4	4
T/Z-Value:	1.25	1.98*	-1.64
Route: 28			
Winter:	16.4/0.07	15.6/0.06	16.6/0.05
Spring:	17.9/0.07	17.5/0.06	18.1/0.01
% Difference:	9	12	9
T/Z-Value:	-5.73*	-8.70*	-3.78*
Route: 31			
Winter:	6.0/0.17	6.1/0.12	6.3/0.04
Spring:	6.5/0.12	6.6/0.10	6.2/0.03
% Difference:	8	8	-2
T/Z-Value:	-2.51*	-3.90*	0.50
Route: 34			
Winter:	7.6/0.14	7.2/0.13	5.6/0.11
Spring:	7.1/0.09	6.7/0.10	5.3/0.07
% Difference:	-7	-7	-5
T/Z-Value:	2.27*	3.23*	0.76
Route: 36			
Winter:	3.4/0.17	3.3/0.13	3.1/0.05
Spring:	3.4/0.14	3.3/0.12	3.2/0.05
% Difference:	0	0	3
T/Z-Value:	0.06	0.63	-1.17
Route: 42			
Winter:	5.6/0.17	6.2/0.14	6.3/0.07
Spring:	5.8/0.23	5.8/0.18	5.8/0.03
% Difference:	4	-6	-8
T/Z-Value:	-0.63	1.64	2.57*

Number of Scheduled Trips between 50 and 100 inclusive

	METHOD 1	METHOD 2	METHOD 3
Route: 29			
Winter:	2.8/0.13	2.9/0.09	3.1/0.15
Spring:	2.3/0.19	2.4/0.14	2.6/0.09
% Difference:	-18	-17	-16
T/Z-Value:	5.75*	7.76*	2.27
Route: 30			
Winter:	2.3/0.23	2.8/0.18	2.2/0.11
Spring:	1.7/0.17	1.9/0.14	2.0/0.05
% Difference:	-26	-32	-9
T/Z-Value:	6.66*	10.29*	1.55
Route: 41			
Winter:	1.4/0.24	2.1/0.18	1.7/0.17
Spring:	1.4/0.21	1.5/0.14	1.5,0.08

% Difference:	0	-29	-12
T/Z-Value:	0.68	8.71*	1.02
Route: 35			
Winter:	2.4/0.20	2.4/0.16	2.4/0.08
Spring:	2.3/0.15	2.3/0.11	2.6/0.04
% Difference:	-4	-4	8
T/Z-Value:	0.50	0.61	-2.55*
Route: 37			
Winter:	1.5/0.25	1.6/0.20	1.7/0.05
Spring:	1.5/0.18	1.5/0.14	1.6/0.09
% Difference:	0	-6	-6
T/Z-Value:	-0.90	1.50	1.76
Route: 24			
Winter:	1.7/0.26	1.7/0.18	2.0/0.14
Spring:	1.8/0.21	1.8/0.15	1.7/0.09
% Difference:	6	6	-15
T/Z-Value:	-1.76	-1.92	2.29
Route: 38			
Winter:	1.1/0.25	1.2/0.20	1.2/0.11
Spring:	1.2/0.24	1.2/0.21	1.2/0.09
% Difference:	9	0	0
T/Z-Value:	-2.34*	-0.39	0.90
Route: 51			
Winter:	2.5/0.13	2.5/0.10	2.6/0.15
Spring:	2.4/0.15	2.4/0.11	2.8/0.05
% Difference:	-4	-4	8
T/Z-Value:	1.67	0.82	-1.01
Route: 26			
Winter:	1.3/0.25	2.5/0.31#	1.4/0.14
Spring:	1.6/0.21	3.8/0.39#	1.3/0.46#
% Difference:	23	52	-7
T/Z-Value:	-3.65*	-5.09*	0.23
Route: 50			
Winter:	1.2/0.21	1.2/0.17	1.2/0.13
Spring:	1.2/0.19	1.2/0.14	1.2/0.11
% Difference:	0	0	0
T/Z-Value:	0.59	-0.23	0.37
Route: 46			
Winter:	1.1/0.25	1.0/0.21	1.1/0.07
Spring:	1.1/0.18	1.0/0.15	1.0/0.15
% Difference:	0	0	-9
T/Z-Value:	0.29	-0.21	1.45

Number of Scheduled Trips < 50

	METHOD 1	METHOD 2	METHOD 3
Route: 52			
Winter:	0.7/0.20	0.7/0.18	0.8/0.37#
Spring:	0.7/0.25	0.7/0.22	0.7/0.19
% Difference:	0	0	-13
T/Z-Value:	0.85	1.62	0.82
Route: 27			
Winter:	0.8/0.27	0.7/0.20	0.8/0.09
Spring:	0.8/0.25	0.8/0.19	0.8/0.07
% Difference:	0	14	0
T/Z-Value:	-0.39	-1.01	0.37

Route: 40			
Winter:	1.4/0.25	1.3/0.19	1.3/0.13
Spring:	1.2/0.20	1.2/0.16	1.3/0.05
% Difference:	-14	-8	0
T/Z-Value:	2.98*	2.70*	0.26
Route: 33			
Winter:	0.6/0.31#	0.6/0.26	0.8/0.20
Spring:	0.6/0.37#	0.6/0.26	0.9/0.65#
% Difference:	0	0	13
T/Z-Value:	0.49	-1.46	-0.37
Route: 14			
Winter:	1.0/0.39#	1.1/0.31#	1.3/0.15
Spring:	1.0/0.33#	1.1/0.22	1.4/0.13
% Difference:	0	0	8
T/Z-Value:	0.04	-0.21	-0.85
Route: 48			
Winter:	0.2/0.32#	0.3/0.28	0.2/0.69#
Spring:	0.2/0.39#	0.2/0.39#	0.1/0.34#
% Difference:	0	-33	-50
T/Z-Value:	3.10*	7.84*	1.41

Note: Each entry in the Winter row or the Spring row contains two items: estimated local bus pass swipe share and coefficient of variation, respectively.

Coefficient of variation (COV) at least 0.3 (rounding effect is considered) or unavailable

* The means are statistically different between the winter and spring of the year (or T value unavailable)

in terms of the likely future effectiveness of each method. Again, please note that Methods 1 and 2 are based on a sample size of 40 for each season whereas Methods 3 and 4 are based on a sample size of 5 for each season. Hence they are not completely comparable.

Table 4.24 is divided into 3 sections. The first section shows the results of the large routes with the number of scheduled trips per day more than 100. The second section shows the results of the medium-sized routes with the number of scheduled trips per day between 50 and 100 inclusive. The last section shows the results of the small routes with the number of scheduled trips per day less than 50. Each entry in the Winter row and the Spring row contains two items. The first item is the estimated local bus pass swipe share and the second item is the corresponding coefficient of variation.

The performance of Methods 1 and 2 is quite similar. Although Method 1 has one fewer route than Method 2 with statistically different mean estimated local bus pass swipe shares between the winter and the spring, Method 2 consistently has lower coefficients of variation (except for route 26) than Method 1. It appears that Method 2 performs slightly better. However, both methods do not appear to perform much better with the large routes than with the small routes. Routes having statistically different mean estimated local bus pass swipe shares between the two seasons are not limited to small or medium-sized routes. Therefore, we can conclude that both methods are not appropriate for estimating daily local bus pass swipes by route.

Method 3 clearly outperforms Methods 1 and 2. The coefficients of variation of Method 3 are generally lower than those of Methods 1 and 2. In addition, fewer routes have statistically different mean estimated local bus pass swipe shares between the winter and the spring. Furthermore, Method 3 works better with large and medium-sized routes than with small routes. Two routes in each of the large and medium-sized categories are considered inappropriate for this method (routes 26, 28, 35, and 42) whereas three small

Table 4.25: Combo Pass Swipes: A Comparison of the Four Methods (1996 data)

1996 Data
Combo Pass Swipe Shares

Number of Scheduled Trips > 100

	METHOD 1	METHOD 2	METHOD 3
Route: 39			
Winter:	15.8/0.09	14.5/0.09	15.5/0.06
Spring:	15.9/0.07	14.8/0.07	15.4/0.04
% Difference:	1	2	-1
T/Z-Value:	-0.29	-1.14	0.30
Route: 32			
Winter:	12.9/0.09	12.4/0.07	13.7/0.08
Spring:	12.6/0.09	12.1/0.08	14.7/0.05
% Difference:	-2	-2	7
T/Z-Value:	1.04	1.70	-1.82
Route: 28			
Winter:	9.8/0.08	9.1/0.07	9.7/0.07
Spring:	10.4/0.09	10.0/0.08	10.2/0.10
% Difference:	6	10	5
T/Z-Value:	-3.33*	-5.36*	-0.98
Route: 31			
Winter:	9.6/0.15	9.6/0.10	9.9/0.08
Spring:	10.2/0.10	10.2/0.08	9.1/0.06
% Difference:	6	6	-8
T/Z-Value:	-1.99*	-2.93*	1.62
Route: 34			
Winter:	11.2/0.11	10.6/0.10	9.4/0.02
Spring:	11.1/0.09	10.3/0.09	9.2/0.09
% Difference:	-1	-3	-2
T/Z-Value:	0.72	1.32	0.66
Route: 36			
Winter:	6.1/0.12	5.9/0.10	5.7/0.09
Spring:	6.4/0.14	6.1/0.12	5.7/0.04
% Difference:	5	3	0
T/Z-Value:	-1.89	-1.79	-0.04
Route: 42			
Winter:	2.2/0.17	2.4/0.14	2.5/0.09
Spring:	2.2/0.27	2.2/0.22	2.3/0.07
% Difference:	0	-8	-8
T/Z-Value:	0.04	2.13*	1.13

Number of Scheduled Trips between 50 and 100 inclusive

	METHOD 1	METHOD 2	METHOD 3
Route: 29			
Winter:	3.2/0.17	3.2/0.15	3.4/0.11
Spring:	2.6/0.18	2.7/0.13	2.8/0.09
% Difference:	-19	-16	-18
T/Z-Value:	5.00*	6.09*	2.84*
Route: 30			
Winter:	2.9/0.25	3.5/0.20	3.0/0.17
Spring:	2.2/0.22	2.4/0.17	2.8/0.15
% Difference:	-24	-31	-7
T/Z-Value:	5.32*	9.12*	0.72
Route: 41			
Winter:	0.8/0.28	1.2/0.21	1.1/0/19
Spring:	0.7/0.26	0.7/0.19	0.8/0.13

% Difference:	-13	-42	-27
T/Z-Value:	4.00*	11.29*	3.29*
Route: 35			
Winter:	3.6/0.19	3.5/0.15	3.4/0.08
Spring:	3.3/0.15	3.3/0.12	3.1/0.08
% Difference:	-8	-6	-9
T/Z-Value:	1.93	2.43*	1.42
Route: 37			
Winter:	2.8/0.22	3.1/0.17	3.3/0.14
Spring:	3.1/0.19	3.0/0.15	3.3/0.20
% Difference:	11	-3	0
T/Z-Value:	-2.00*	0.34	-0.03
Route: 24			
Winter:	1.7/0.24	1.7/0.17	1.7/0.07
Spring:	2.0/0.19	1.9/0.14	1.7/0.11
% Difference:	18	12	0
T/Z-Value:	-3.26*	-3.90*	-0.51
Route: 38			
Winter:	1.6/0.25	1.8/0.19	1.8/0.12
Spring:	1.8/0.22	1.8/0.19	2.0/0.15
% Difference:	13	0	11
T/Z-Value:	-2.30*	-0.13	-1.31
Route: 51			
Winter:	4.3/0.14	4.1/0.11	4.5/0.11
Spring:	3.9/0.16	3.8/0.11	4.7/0.03
% Difference:	-9	-7	4
T/Z-Value:	3.03*	2.96*	-0.76
Route: 26			
Winter:	2.7/0.30#	4.8/0.31#	3.5/0.35#
Spring:	3.3/0.22	6.4/0.26	3.0/0.54#
% Difference:	22	33	-14
T/Z-Value:	-3.55*	-4.66*	0.53
Route: 50			
Winter:	2.1/0.20	2.1/0.17	2.2/0.10
Spring:	2.0/0.17	2.0/0.13	2.3/0.10
% Difference:	-5	-5	5
T/Z-Value:	1.29	0.79	-0.42
Route: 46			
Winter:	0.5/0.26	0.4/0.23	0.5/0.16
Spring:	0.4/0.23	0.4/0.21	0.5/0.27
% Difference:	-20	0	0
T/Z-Value:	1.28	0.91	0.18

Number of Scheduled Trips < 50

	METHOD 1	METHOD 2	METHOD 3
Route: 52			
Winter:	1.6/0.15	1.6/0.14	1.5/0.24
Spring:	1.7/0.23	1.7/0.21	1.6/0.23
% Difference:	6	6	7
T/Z-Value:	-1.17	-0.59	-0.33
Route: 27			
Winter:	1.3/0.26	1.2/0.21	1.3/0.13
Spring:	1.3/0.23	1.3/0.18	1.3/0.32#
% Difference:	0	8	0
T/Z-Value:	-0.16	-0.62	0.08

Route: 40			
Winter:	2.2/0.22	2.1/0.17	2.0/0.10
Spring:	1.8/0.19	1.8/0.15	2.0/0.09
% Difference:	-18	-14	0
T/Z-Value:	3.86*	3.65*	-0.05
Route: 33			
Winter:	0.6/0.43#	0.5/0.40#	0.7/0.16
Spring:	0.6/0.37#	0.6/0.26	1.0/1.16#
% Difference:	0	20	43
T/Z-Value:	0.10	-1.35	-0.59
Route: 14			
Winter:	0.4/0.43#	0.4/0.33#	0.4/0.45#
Spring:	0.4/0.40#	0.4/0.30#	0.4/0.12
% Difference:	0	0	0
T/Z-Value:	-0.85	-1.06	-0.32
Route: 48			
Winter:	0.1/0.38#	0.2/0.27	0.1/1.27#
Spring:	0.1/0.48#	0.1/0.45#	0.1/0.89#
% Difference:	0	-50	0
T/Z-Value:	1.02	5.85*	1.05

Note: Each entry in the Winter row or the Spring row contains two items: estimated combo pass swipe share and coefficient of variation, respectively.

Coefficient of variation (COV) at least 0.3 (rounding effect is considered) or unavailable

* The means are statistically different between the winter and spring of the year (or T value unavailable)

routes have high coefficients of variation (routes 33, 48, and 52). However, as mentioned before, if the sample size is increased, it is possible that the coefficients of variation of these small routes can be reduced. Finally, the data in 1996 are so good that all routes have sufficient data remaining after all eliminations, which makes Method 4 unnecessary. This is the reason why Method 4 is not performed for pass swipe analyses.

Therefore, we can conclude that Method 3 is the best method among the three proposed for estimating daily local bus pass swipes by route.

Table 4.25 shows the comparison of estimated combo pass swipe shares, coefficients of variation, percentage difference in estimated combo pass swipe shares for the two seasons and t/z values produced by the three different methods based on the 1996 data.

As in the local bus pass swipe analysis, Method 2 produces lower coefficients of variation for most routes, but slightly more routes with statistically different mean estimated combo bus pass swipe shares than Method 1 does. Therefore, Method 2 appears to be slightly better than Method 1. Also, routes having statistically different estimated combo pass swipe shares for the two seasons include routes of all sizes; therefore, we conclude that Methods 1 and 2 are not satisfactory in estimating daily combo pass swipes by route.

Method 3 clearly outperforms the other two methods, at least for the large and medium-sized routes. First, in most cases Method 3 produces lower coefficients of variation than the other two methods do. Second, Method 3 produces fewer large and medium-sized routes in having statistically different mean estimated combo bus pass swipe shares between the two seasons than the other two methods do. In fact, no large route and only two medium-sized routes have statistically different mean estimated combo bus pass swipes between the two seasons for Method 3 compared to a total of 10 large and medium-sized routes for each of Method 1 and Method 2.

Therefore, we can conclude that Method 3 is the best method among the three proposed for estimating daily combo pass swipes by route.

4.4 Daily Ridership by Route

This section presents the results of estimating daily route ridership. The daily route ridership is simply the sum of the daily pass ridership, as estimated in Section 4.3, and the daily fare ridership. Only one method is used in this analysis. As mentioned in Chapter 3, since the unknown ridership given in the MBTA Revenue and Ridership Summary by Route is not very accurate, and that there is no obvious control total for ridership, the concept of average fare is used to estimate the daily fare ridership by route based on total revenue per route estimated in Section 4.2. Since the spring 1996 average fares are clearly better than the pooled average fares, all analyses in this thesis uses the spring 1996 average fares. The ridership estimates presented in this section and in the following section are route ridership, which contains all fare and pass ridership. Since the method relies heavily on the screening part of Method 3 for estimating daily fare revenue by route, it is referred to as Method 3 here. This is to remind the reader that trip-level screening is required in this method.

Table 4.26 shows the estimated daily route ridership for each route with the corresponding daily fare and pass ridership based on combined winter and spring 1996 data. The table shows that about 57% to 79% of ridership is fare ridership, and the remaining ridership is pass ridership.

Table 4.27 shows the results of the analysis.

Table 4.26: Daily Fare and Pass Ridership by Route (Combined 1996 Data)

Route	Ridership	Fare Ridership (%)	Pass Ridership (%)
14	570	355 (62)	215 (38)
24	1624	1222 (75)	402 (25)
26	1716	1247 (73)	469 (27)
27	610	396 (65)	214 (35)
28	9163	6151 (67)	3013 (33)
29	1618	1041 (64)	577 (36)
30	1521	989 (65)	532 (35)
31	4210	2645 (63)	1564 (37)
32	6680	4346 (65)	2334 (35)
33	590	413 (70)	177 (30)
34	2731	1640 (60)	1092 (40)
34E	1921	1362 (71)	559 (29)
35	1922	1316 (68)	606 (32)
36	2627	1673 (64)	954 (36)
37	1640	1115 (68)	525 (32)
38	745	439 (59)	306 (41)
39	16698	9487 (57)	7211 (43)
40	951	619 (65)	332 (35)
41	955	651 (68)	304 (32)
42	3084	2096 (68)	988 (32)
46	459	283 (62)	177 (39)
48	97	77 (79)	20 (21)
50	897	550 (61)	347 (39)
51	722	508 (70)	215 (30)
52	1921	1362 (71)	559 (29)

In 1994, only 5 out of 25 routes (routes 14, 26, 30, 33, and 48) have coefficients of variation greater than 0.3, and they are all medium-sized or small routes with the estimated number of riders no more than 2000. The results are more impressive in 1996 since only 2 routes have coefficients of variation greater than 0.3; the loop routes 26 and 48. Therefore, we can see that this method provides stable data within a season. Unfortunately, 36% of the routes in 1994 and 32% of the routes in 1996 have the mean estimated daily ridership by route in the winter statistically different from that of the spring, and this set includes routes of all sizes. However, the statistically different means of the large routes are probably due to the very low coefficients of variation these routes have; typically below 0.10.

Table 4.28 shows a comparison of the daily ridership estimates provided by Method 3 and the daily ridership information collected from ride checks for routes 39 and 42. Since the ride checks were conducted in the winter of 1996, only ridership estimates from the same season are used. Unfortunately, ride check ridership information for other routes or other time periods is not available.

Table 4.28 shows that the electronic farebox daily ridership estimates are lower than the ridership information collected from ride checks for both routes 39 and 42. The electronic farebox underestimated the daily ridership of routes 39 and 42 by 15% and 2%, respectively. The estimate for route 42 is excellent particularly in light of the small sample size for the electronic farebox estimates and the measurement and sampling error for the ride check estimates, but the one for route 39 is less encouraging. However, route 39 is a complex route, and there are two possible explanations for the underestimation of daily ridership by the electronic farebox for route 39. First, route 39 is operated primarily to substitute for the suspended portion of the Green Line (E Line) between Heath Street and

Table 4.27: Daily Ridership by Route: Method 3 Results

Rt	Winter Mean	COV	1994 Spring Mean	COV	T	Winter Mean	COV	1996 Spring Mean	COV	T
39	14194	0.09	16673	0.05	-3.53*	16523	0.04	16873	0.05	-0.75
28	9478	0.04	9998	0.06	-1.60	8332	0.09	9995	0.04	-4.43*
32	6597	0.06	6711	0.06	-0.46	6227	0.08	7133	0.03	-3.68*
31	4275	0.08	4064	0.09	0.97	4075	0.09	4344	0.05	-1.51
34	2781	0.12	3066	0.11	-1.32	2656	0.06	2806	0.06	-1.46
42	2522	0.15	3247	0.04	-3.98*	3287	0.06	2882	0.06	3.64*
36	2350	0.04	2811	0.09	-3.97*	2529	0.03	2724	0.07	-2.17
34E	2590	0.16	3087	0.10	-2.18	1999	0.19	1844	0.09	0.83
35	2580	0.15	1781	0.18	3.54*	1725	0.15	2119	0.05	-3.07*
29	1945	0.11	2241	0.15	-1.65	1592	0.21	1645	0.02	-0.34
51	1607	0.09	1817	0.13	-1.75	1653	0.11	1805	0.04	-1.71
37	1784	0.08	1592	0.09	2.13	1668	0.11	1611	0.14	0.44
26	1581	0.40#	1908	1.05#	-0.35	1146	NA#	1830	0.30#	NA*
30	1122	0.32#	1513	0.11	-2.20	1561	0.17	1481	0.10	0.60
24	1686	0.14	1328	0.08	3.03*	1677	0.08	1571	0.05	1.57
40	827	0.11	1057	0.19	-2.35*	911	0.13	991	0.10	-1.14
50	895	0.27	891	0.05	0.04	811	0.05	982	0.12	-3.22*
41	658	0.23	643	0.11	0.22	1026	0.26	884	0.07	1.16
38	926	0.11	805	0.07	2.24	646	0.16	843	0.09	-3.45*
52	844	0.21	747	0.11	1.13	706	0.15	739	0.21	-0.39
27	680	0.28	550	0.13	1.43	608	0.10	613	0.15	-0.10
14	NA	NA#	433	0.29	NA*	532	0.10	609	0.10	-2.18
33	656	NA#	574	0.11	NA*	598	0.14	582	0.28	0.20
46	484	0.14	431	0.18	1.15	454	0.09	464	0.06	-0.50
48	NA	NA#	137	0.42#	NA*	NA	NA#	97	0.42#	NA*
T	63062	0.14	68105	0.11		62942	0.11	67467	0.07	

* The means are statistically different between the winter and spring of the year (or T value unavailable)

Coefficient of variation (COV) at least 0.3 (or unavailable)

T Total mean estimated ridership for the garage and median COV

Table 4.28: A Comparison of Electronic Farebox Daily Ridership Estimates and Ride Check Daily Ridership Information (Winter 1996)

Route	Ride Check Daily Ridership Information	Electronic Farebox Daily Ridership Estimate	Percentage Difference
39	19425	16523	-15
42	3340	3287	-2

Arborway. Therefore, transfers are widely used. Some drivers may not register transfers with the electronic farebox, which would mean that these transfers are not counted, leading to a higher estimated average fare, and hence a lower estimated ridership. Second, since route 39 is a heavy route, occasionally back door boarding may be practiced, and in this case, pass riders and transfer riders will not be counted. Again, this leads to a higher estimated average fare and a lower estimated ridership. Since both routes 39 and 42 are large routes, we can expect the electronic farebox daily ridership estimates for the small routes to be less accurate.

From the above discussion, we can see that Method 3 generates low coefficients of variation in general. Unfortunately, quite a few routes have statistically different mean estimated daily route ridership between the winter and spring of each year. In addition, since ride check data are only available for two routes, it is difficult to conclude if the ridership estimates are reliable. The ridership estimate for route 42 is encouraging, but the ridership estimate for route 39 is less encouraging.

4.5 Ridership by Route and Time Period

There is also only one method for estimating ridership by route and time period. In fact, this method relies on the one for estimating daily ridership by route described in the previous section. Therefore, it is also named Method 3 here to remind the reader that trip-level screening is required. As described in Chapter 3, the proposed method consists of two parts. The first part involves estimating the ridership of each route for the day of interest by applying the method proposed in Section 3.5. The second part involves allocating the ridership to the five time periods of the day with the daily ridership serving as the control total. The second part attempts to correct two types of errors: improper fare rider registration for transfer riders and riders who board from the back door, and missing trip records.

Table 4.29 through 4.33 show the total number of scheduled trips, the total number of trip records, and the total number of trips remaining after the screening for the ten days analyzed for all five periods based on 1996 data. Table 4.29 shows that for some routes, there are more trip records than the number of scheduled trips for this time period. In fact, the number of trip records in this period is 89% of the number of scheduled trips, which is the highest among the five time periods. One possible explanation for this is that the electronic farebox automatically generates a trip record at midnight of each day even when there is no scheduled trip at midnight for the route. However, these "extra" trip records are eliminated by the screening procedure, since the number of remaining records is only 50% of the number of scheduled trips, which is comparable to other time periods. For the other four time periods, the results are quite similar. The number of trip records is about 66% to 78% of the number of scheduled trips, and the number of remaining records is about 41% to 52% of the number of scheduled trips.

Table 4.29: Ridership by Route and Time Period: Screening Results (Early Morning)

Route	Sch_Trips	Records (% of Sch)	Remaining (% of Sch)
14	30	21 (70)	16 (53)
24	100	97 (97)	63 (63)
26	100	29 (28)	4 (4)
27	70	81 (115)	68 (97)
28	270	307 (113)	142 (52)
29	120	133 (110)	54 (45)
30	120	113 (94)	69 (57)
31	280	209 (74)	121 (43)
32	370	336 (90)	229 (61)
33	30	28 (93)	14 (46)
34	330	214 (64)	121 (36)
35	100	88 (88)	60 (60)
36	330	247 (74)	126 (38)
37	90	58 (64)	34 (37)
38	90	59 (65)	38 (42)
39	440	532 (120)	293 (66)
40	30	55 (183)	42 (140)
41	150	40 (26)	30 (20)
42	130	107 (82)	56 (43)
46	10	19 (190)	13 (130)
48	0	0 (0)	0 (0)
50	30	31 (103)	22 (73)
51	60	61 (101)	39 (65)
52	20	35 (175)	16 (80)
340	120	156 (130)	71 (59)
Total	3420	3056 (89)	1741 (50)

Table 4.30: Ridership by Route and Time Period: Screening Results (AM Peak)

Route	Sch_Trips	Records (% of Sch)	Remaining (% of Sch)
14	40	25 (62)	22 (55)
24	120	102 (85)	92 (76)
26	80	7 (8)	1 (1)
27	80	78 (97)	73 (91)
28	380	332 (87)	212 (55)
29	150	83 (55)	66 (44)
30	120	43 (35)	35 (29)
31	300	178 (59)	145 (48)
32	540	299 (55)	208 (38)
33	40	29 (72)	22 (55)
34	230	179 (77)	117 (50)
35	130	68 (52)	46 (35)
36	240	162 (67)	97 (40)
37	130	48 (36)	35 (26)
38	100	57 (56)	42 (42)
39	710	528 (74)	346 (48)
40	80	71 (88)	43 (53)
41	140	26 (18)	19 (13)
42	230	139 (60)	91 (39)
46	80	79 (98)	78 (97)
48	0	7 (0)	2 (0)
50	100	70 (70)	56 (56)
51	110	80 (72)	68 (61)
52	100	50 (50)	33 (33)
340	140	159 (113)	73 (52)
Total	4370	2899 (66)	2022 (46)

Table 4.31: Ridership by Route and Time Period: Screening Results (Base)

Route	Sch_Trips	Records (% of Sch)	Remaining (% of Sch)
14	140	60 (42)	46 (32)
24	230	144 (62)	67 (29)
26	190	61 (32)	12 (6)
27	160	159 (99)	132 (82)
28	770	632 (82)	337 (43)
29	180	114 (63)	37 (20)
30	360	214 (59)	117 (32)
31	740	569 (76)	428 (57)
32	710	595 (83)	407 (57)
33	140	101 (72)	58 (41)
34	330	328 (99)	191 (57)
35	350	267 (76)	183 (52)
36	330	261 (79)	179 (54)
37	300	243 (81)	150 (50)
38	260	161 (61)	102 (39)
39	1340	1224 (91)	807 (60)
40	200	168 (84)	124 (62)
41	290	206 (71)	140 (48)
42	500	327 (65)	187 (37)
46	280	244 (87)	199 (71)
48	140	58 (41)	22 (15)
50	240	192 (80)	126 (52)
51	230	170 (73)	119 (51)
52	220	189 (85)	105 (47)
340	280	345 (123)	155 (55)
Total	8910	7032 (78)	4430 (49)

Table 4.32: Ridership by Route and Time Period: Screening Results (PM Peak)

Route	Sch_Trips	Records (% of Sch)	Remaining (% of Sch)
14	40	40 (100)	37 (92)
24	80	62 (77)	43 (53)
26	80	13 (16)	2 (2)
27	80	45 (56)	37 (46)
28	380	262 (68)	178 (46)
29	150	89 (59)	63 (42)
30	110	6 (5)	5 (4)
31	300	174 (57)	131 (43)
32	380	307 (80)	260 (68)
33	40	25 (62)	16 (40)
34	240	175 (72)	131 (54)
35	160	116 (72)	76 (47)
36	160	124 (77)	95 (59)
37	160	121 (75)	100 (62)
38	100	84 (84)	68 (68)
39	540	471 (87)	284 (52)
40	80	63 (78)	53 (66)
41	120	74 (61)	55 (45)
42	120	95 (79)	71 (59)
46	80	80 (100)	72 (90)
48	10	0 (0)	0 (0)
50	120	86 (71)	75 (62)
51	120	82 (68)	76 (63)
52	70	39 (55)	29 (41)
340	110	150 (136)	67 (60)
Total	3830	2783 (72)	2024 (52)

Table 4.33: Ridership by Route and Time Period: Screening Results (Evening)

Route	Sch_Trips	Records (% of Sch)	Remaining (% of Sch)
14	20	20 (100)	9 (45)
24	180	123 (68)	70 (38)
26	130	18 (13)	5 (3)
27	50	22 (44)	14 (28)
28	550	397 (72)	245 (44)
29	290	234 (80)	147 (50)
30	170	82 (48)	55 (32)
31	420	293 (69)	183 (43)
32	410	329 (80)	195 (47)
33	30	15 (50)	9 (30)
34	250	154 (61)	98 (39)
35	100	55 (55)	27 (27)
36	300	236 (78)	140 (46)
37	120	80 (66)	37 (30)
38	110	73 (66)	37 (33)
39	980	727 (74)	443 (45)
40	50	19 (38)	3 (6)
41	170	83 (48)	67 (39)
42	210	135 (64)	76 (36)
46	60	48 (80)	32 (53)
48	0	0 (0)	0 (0)
50	70	57 (81)	28 (40)
51	120	80 (66)	51 (42)
52	60	30 (50)	28 (46)
340	170	195 (114)	88 (51)
Total	5020	3505 (69)	2087 (41)

Tables 4.34 through 4.38 show the overall results for 5 time periods in each day: Early Morning (before 7 am), AM Peak (7 - 9 am), Base (9 am - 4 pm), PM Peak (4 - 6 pm), and Evening (after 6 pm).

Before looking at the results, it is important to understand that the estimation process for each season depends on the availability of data for all five time periods. Since the estimation process is basically the allocation of the estimated daily route ridership to the five time periods, if a route has one or more time period without data for all five days sampled in a season, all other estimated route ridership by time period would be biased (inflated) because these estimates would include the ridership that should be allocated to the time periods without data. Therefore, in order to avoid bias, if a route has no data (among the five days sampled) for one or more time period, it will not be analyzed for that season. The mean, coefficient of variation, and t-value of these routes are labelled "NA" in Tables 4.34 through 4.38. For 1994, routes 14, 26, 27, 33, 40, and 48 were not analyzed in both seasons; routes 35 and 41 were not analyzed in the winter; and route 52 was not analyzed in the spring. For 1996, route 48 was not analyzed for both seasons, and routes 26, 30, 37, 40, and 41 were not analyzed in the winter. We can see that the spring 1996 data are of higher quality than those of the other seasons.

From Table 4.34, we can see that for the Early Morning period, only routes 34 and 51 give totally satisfactory results in 1994. In 1996, a total of nine routes (routes 24, 27, 28, 31, 34, 36, 39, 46, and 51) give totally satisfactory results: all relatively large routes except route 46. However, not all large routes give satisfactory results. Some small routes have no data for one (or more) seasons because not many trips are scheduled for these routes for this period. This is especially common in 1994 and winter 1996. From the substantial improvement of the results between 1994 and 1996, we may conclude that if driver compliance continues to improve, there is potential that the method may be used to

Table 4.34: Ridership by Route and Time Period: Method 3 Results (Early Morning)

Rt	Winter Mean	COV	1994 Spring Mean	COV	T	Winter Mean	COV	1996 Spring Mean	COV	T
39	1037	0.07	1257	0.09	-3.58*	1425	0.11	1368	0.11	0.58
28	1051	0.25	855	0.30#	1.18	1234	0.21	1079	0.27	0.89
32	605	0.50#	558	0.07	0.34	627	0.12	1031	0.09	-7.76*
34	574	0.21	533	0.29	0.47	639	0.16	665	0.22	-0.33
31	278	0.99#	514	0.39#	-1.54	681	0.29	629	0.07	0.58
36	309	0.31#	478	0.53#	-1.39	453	0.09	494	0.13	-1.21
42	206	0.37#	380	0.21	-3.52*	494	0.63#	261	0.15	1.65
24	294	0.19	233	0.36#	1.35	368	0.10	369	0.17	-0.04
34E	328	0.49#	360	0.14	-0.42	89	0.31#	206	0.13	-6.74*
51	185	0.16	159	0.28	1.08	207	0.23	201	0.05	0.30
30	109	0.69#	165	0.37#	-1.28	NA	NA#	225	0.37#	NA*
37	116	0.25	181	0.65#	-1.19	NA	NA#	179	0.31#	NA*
35	NA	NA#	151	0.34#	NA*	123	0.22	181	0.08	-4.28*
38	169	0.30#	146	NA#	NA*	94	0.72#	138	0.18	-1.35
27	NA	NA#	NA	NA#	NA*	118	0.21	140	0.20	-1.30
26	NA	NA#	NA	NA#	NA*	NA	NA#	98	NA#	NA*
33	NA	NA#	NA	NA#	NA*	96	0.09	82	0.69#	0.55
29	95	0.77#	44	1.55#	1.14	26	0.69#	168	0.09	-13.49*
40	NA	NA#	NA	NA#	NA*	NA	NA#	80	0.20	NA*
14	NA	NA#	NA	NA#	NA*	55	0.22	72	0.10	-2.80*
41	NA	NA#	42	0.41#	NA*	NA	NA#	73	0.27	NA*
50	62	0.89#	65	0.50#	-0.09	54	0.23	21	0.40#	4.84*
52	43	0.46#	NA	NA#	NA*	8	1.84#	53	0.12	-5.76*
46	15	0.45#	14	0.55#	0.22	16	0.27	13	0.12	1.05
T	5476	0.50#	6135	0.46#		6807	0.25	7826	0.16	

* The means are statistically different between the winter and spring of the year (or T value unavailable)

\# Coefficient of variation (COV) at least 0.3 (or unavailable)

T Total mean estimated ridership for the garage and median COV

Table 4.35: Ridership by Route and Time Period: Method 3 Results (AM Peak)

Rt	1994 Winter Mean	COV	1994 Spring Mean	COV	T	1996 Winter Mean	COV	1996 Spring Mean	COV	T
39	3235	0.28	3394	0.11	-0.36	3285	0.16	3519	0.13	-0.75
32	1526	0.13	1758	0.35#	-0.80	1494	0.09	1732	0.03	-3.63*
28	1360	0.12	1452	0.15	-0.74	1046	0.10	1395	0.08	-4.89*
42	554	0.17	507	0.15	0.89	657	0.18	683	0.09	-0.43
31	620	0.22	513	0.20	1.40	653	0.18	572	0.11	1.35
36	555	0.14	277	0.98#	2.19	621	0.04	786	0.25	-1.88
34	522	0.20	533	0.15	-0.18	560	0.24	392	0.13	2.64*
34E	475	0.21	422	0.13	1.05	373	0.26	328	0.19	0.87
37	552	0.19	313	NA#	NA*	NA	NA#	283	0.15	NA*
51	348	0.06	391	0.15	-1.56	341	0.13	407	0.09	-2.56*
29	361	0.13	646	NA#	NA*	141	0.30	248	0.06	-5.39*
24	298	0.19	277	0.18	0.62	436	0.34#	333	0.05	1.54
30	172	0.74#	247	0.24	-1.19	NA	NA#	405	0.47#	NA*
52	266	0.25	NA	NA#	NA*	302	0.71#	202	0.11	1.02
35	NA	NA#	182	0.28	NA*	257	0.31#	307	0.14	-1.22
41	NA	NA#	117	0.35#	1.17	NA	NA#	261	0.22	NA*
40	NA	NA#	NA	NA#	NA*	NA	NA#	189	0.21	NA*
50	186	0.35#	150	0.60#	0.71	155	0.21	184	0.14	-1.54
38	184	0.19	103	1.19#	1.42	155	0.20	183	0.24	-1.16
27	NA	NA#	NA	NA#	NA*	114	0.11	132	0.29	-1.02
33	NA	NA#	NA	NA#	NA*	127	0.10	101	0.21	2.40*
26	NA	NA#	NA	NA#	NA*	NA	NA#	107	NA#	NA*
46	78	0.38#	85	0.10	-0.50	89	0.21	64	0.19	2.45*
14	NA	NA#	NA	NA#	NA*	70	0.49#	85	0.22	-0.82
T	11292	0.24	11367	0.35#		10876	0.23	12898	0.15	

* The means are statistically different between the winter and spring of the year (or T value unavailable)

Coefficient of variation (COV) at least 0.3 (or unavailable)

T Total mean estimated ridership for the garage and median COV

Table 4.36: Ridership by Route and Time Period: Method 3 Results (Base)

Rt	1994 Winter Mean	COV	1994 Spring Mean	COV	T	1996 Winter Mean	COV	1996 Spring Mean	COV	T
39	5700	0.12	6335	0.03	-2.10	6138	0.05	6326	0.03	-1.23
28	3844	0.07	3753	0.10	0.44	3634	0.17	4351	0.06	-2.36*
32	2361	0.07	2267	0.07	0.87	2264	0.14	2493	0.07	-1.46
31	1917	0.13	1451	0.06	3.98*	1541	0.14	1620	0.06	-0.72
42	1198	0.23	1509	0.06	-2.42*	1463	0.18	1243	0.12	1.63
34E	963	0.17	1460	0.14	-4.19*	1026	0.34#	711	0.07	2.02
35	NA	NA#	925	0.20	NA*	928	0.27	1136	0.07	-1.75
36	806	0.13	1086	0.28	-1.91	859	0.16	727	0.08	2.00
34	674	0.44#	835	0.16	-1.09	659	0.15	852	0.12	-3.01*
26	NA	NA#	NA	NA#	NA*	NA	NA#	691	0.15	NA*
37	615	0.24	650	0.11	-0.47	NA	NA#	673	0.11	NA*
51	556	0.15	617	0.17	-1.00	644	0.14	603	0.04	1.02
30	453	0.51#	681	0.23	-1.81	NA	NA#	423	0.33#	NA*
29	548	0.35#	616	0.21	-0.66	520	0.37#	315	0.14	2.30
40	NA	NA#	NA	NA#	NA*	NA	NA#	479	0.17	NA*
24	663	0.36#	394	0.21	2.39*	360	0.10	481	0.13	-3.68*
50	378	0.39#	377	0.11	0.01	369	0.12	469	0.12	-3.18*
41	NA	NA#	337	0.05	NA*	NA	NA#	369	0.08	NA*
52	402	0.26	NA	NA#	NA*	277	0.68#	314	0.34#	-0.39
14	NA	NA#	NA	NA#	NA*	297	0.31#	359	0.16	-1.29
38	357	0.14	338	0.13	0.64	212	0.19	356	0.07	-6.69*
33	NA	NA#	NA	NA#	NA*	281	0.24	261	0.20	0.54
46	292	0.09	241	0.27	1.66	271	0.15	278	0.08	-0.29
27	NA	NA#	NA	NA#	NA*	233	0.13	183	0.17	2.49*
48	NA	NA#	NA	NA#	NA*	NA	NA#	NA	NA#	NA*
T	21727	0.35#	23872	0.20		21976	0.19	25713	0.12	

* The means are statistically different between the winter and spring of the year (or T value unavailable)

Coefficient of variation (COV) at least 0.3 (or unavailable)

T Total mean estimated ridership for the garage and median COV

Table 4.37: Ridership by Route and Time Period: Method 3 Results (PM Peak)

Rt	1994 Winter Mean	COV	1994 Spring Mean	COV	T	1996 Winter Mean	COV	1996 Spring Mean	COV	T
39	2122	0.13	2773	0.12	-3.42*	2805	0.07	2619	0.08	1.43
28	1710	0.23	1870	0.16	-0.72	1220	0.18	1559	0.08	-2.92*
32	1269	0.18	1169	0.21	0.67	1052	0.13	956	0.02	1.59
31	755	0.32#	914	0.10	-1.37	672	0.48#	690	0.06	-0.12
34	579	0.28	583	0.16	-0.04	435	0.25	472	0.16	-0.64
26	NA	NA#	NA	NA#	NA*	NA	NA#	450	NA#	NA*
29	465	0.26	399	0.22	0.97	415	0.14	418	0.13	-0.07
42	330	0.25	424	0.13	-2.11	384	0.26	365	0.14	0.37
35	NA	NA#	397	0.32#	NA*	309	0.25	369	0.16	-1.35
51	308	0.29	362	0.14	-1.18	281	0.12	373	0.06	-5.10*
37	331	0.20	274	0.23	1.39	NA	NA#	315	0.17	NA*
36	281	0.21	330	0.35#	-0.84	199	0.40#	314	0.19	-2.57*
34E	365	0.33#	327	0.15	0.64	107	0.56#	280	0.08	-5.98*
30	225	0.25	216	0.23	0.26	NA	NA#	278	NA#	NA*
24	193	0.37#	216	0.30	-0.52	187	0.18	181	0.18	0.27
50	188	0.12	182	0.12	0.46	149	0.19	206	0.18	-2.73*
40	NA	NA#	NA	NA#	NA*	NA	NA#	149	0.16	NA*
38	144	0.26	149	0.17	-0.24	98	0.42#	104	0.29	-0.26
27	NA	NA#	NA	NA#	NA*	96	0.17	119	0.27	-1.45
52	100	0.16	NA	NA#	NA*	97	0.70#	120	0.17	-0.73
41	NA	NA#	88	NA#	NA*	NA	NA#	121	0.12	NA*
14	NA	NA#	NA	NA#	NA*	81	0.26	72	0.34#	0.61
33	NA	NA#	NA	NA#	NA*	44	0.59#	101	0.36#	-2.84*
46	71	0.28	71	0.19	0.02	58	0.12	76	0.17	-2.70*
48	NA	NA#	NA	NA#	NA*	NA	NA#	NA	NA#	NA*
T	9436	0.28	10744	0.23		8689	0.26	10707	0.17	

* The means are statistically different between the winter and spring of the year (or T value unavailable)

Coefficient of variation (COV) at least 0.3 (or unavailable)

T Total mean estimated ridership for the garage and median COV

Table 4.38: Ridership by Route and Time Period: Method 3 Results (Evening)

Rt	1994 Winter Mean	COV	1994 Spring Mean	COV	T	1996 Winter Mean	COV	1996 Spring Mean	COV	T
39	2101	0.08	2914	0.07	-6.85*	2870	0.07	3040	0.07	-1.33
27	NA	NA#	NA	NA#	NA*	1199	0.18	1611	0.10	-3.46*
28	1514	0.22	2069	0.19	-2.39*	488	0.27	494	0.11	-0.10
31	705	0.13	672	0.20	0.45	790	0.17	921	0.10	-1.79
32	836	0.14	959	0.20	-1.21	48	NA#	37	0.33#	NA*
36	396	0.13	640	0.37#	-2.26	395	0.30#	401	0.14	-0.10
34	430	0.28	581	0.31#	-1.57	363	0.21	425	0.26	-1.04
34E	457	0.39#	518	0.21	-0.65	402	0.27	319	0.20	1.48
30	162	0.28	203	0.19	-1.54	NA	NA#	833	0.10	-5.16*
29	474	0.33#	533	0.17	-0.73	133	0.40#	148	0.44#	-0.37
42	234	0.22	427	0.21	-4.22*	286	0.31#	328	0.24	-0.80
33	NA	NA#	NA	NA#	NA*	83	NA#	482	NA#	NA*
51	210	0.19	286	0.47#	-1.22	178	0.25	220	0.17	-1.64
26	NA	NA#	NA	NA#	NA*	NA	NA#	207	0.08	NA*
37	167	0.39#	171	0.42#	-0.10	NA	NA#	161	0.30	NA*
40	207	NA#	NA	NA#	NA*	NA	NA#	92	0.30#	NA*
24	237	0.46#	207	0.20	0.56	47	NA#	37	0.19	NA*
35	NA	NA#	126	0.25	NA*	107	0.24	125	0.36#	-0.78
50	80	0.46#	115	0.81#	-0.78	82	0.60#	102	0.33#	-0.71
38	71	0.44#	69	0.41#	0.12	85	0.36#	60	0.24	1.67
41	NA	NA#	59	NA#	NA*	NA	NA#	59	0.20	NA*
52	31	0.40#	NA	NA#	NA*	20	0.93#	47	0.38#	-2.25
46	26	0.35#	20	0.11	1.47	19	0.20	32	0.26	-3.22*
14	NA	NA#	NA	NA#	NA*	28	0.57#	19	0.36#	1.14
T	8338	0.39#	10569	0.34#		7623	0.38#	10200	0.24	

* The means are statistically different between the winter and spring of the year (or T value unavailable)

\# Coefficient of variation (COV) at least 0.3 (or unavailable)

T Total mean estimated ridership for the garage and median COV

estimate ridership by route and time period for the Early Morning period for the large routes, however a larger sample size than 5 days per season would certainly be required for the smaller routes.

The AM Peak period has more encouraging results. In 1994, 8 out of 24 routes (routes 24, 28, 31, 34, 34E, 39, 42, and 51) are totally satisfactory; all relatively large routes. In fact, many small routes have no data remaining after the screening because there is little data to start with. In 1996, 8 routes (routes 27, 31, 34E, 36, 38, 39, 42, and 50) give satisfactory results, including routes of all sizes. Although not all large routes are considered satisfactory, if driver compliance continues to improve, it is likely that the method can be used to estimate ridership by route and time period for the large routes for the AM Peak period. A larger sample size will be required for smaller routes.

A total of 8 routes in 1994 and 9 routes in 1996 are considered satisfactory in the Base period analysis. This includes routes of all sizes, but most of them are relatively large routes. Therefore, it is likely that with continued driver compliance improvement, this method can estimate ridership by route by time period for the large routes for the Base period.

For the PM Peak period, 11 routes in 1994 and 8 routes in 1996 have satisfactory results. Again, they are generally relatively large routes. Therefore, as with the Base period, it is quite likely that this method can be used to estimate ridership by route and time period for the large routes for the PM Peak period. Again, for smaller routes a larger sample size will be required.

The results of the Evening period are very similar to those of the Early Morning period. Only 3 routes in 1994 and 6 routes in 1996 have satisfactory results, and they are all relatively large routes. However, other large routes do not have satisfactory results. Therefore, again, it is hard to estimate ridership for the Evening period by this method at this time, but with driver compliance improvement, there is some potential that the estimation could be satisfactorily made for the larger routes.

Table 4.39 compares ride check ridership information by time period and electronic farebox ridership by time period estimates.

Table 4.39: A Comparison of Electronic Farebox Ridership by Time Period Estimates and Ride Check Ridership by Time Period Information (Winter 1996)

			Route 39			Route 42	
		RC	EF	%Diff	RC	EF	%Diff
Early Morning	(12 AM - 7 AM)	1256	1425	13	223	494	122
AM Peak	(7 AM - 9 AM)	3906	3285	-16	759	657	-13
Base	(9 AM - 4 PM)	7643	6138	-20	1638	1463	-11
PM Peak	(4 PM - 6 PM)	3137	2805	-11	437	384	-12
Evening	(6 PM - 12 AM)	3483	2870	-18	283	286	1
Total		19425	16523	-15	3340	3284	-2

Table 4.39 shows that for route 39, PM Peak, Early Morning, and AM Peak periods have the best results whereas for route 42, the Evening period has the best results, followed by the Base and the two Peak periods. Although the good results of the Early Morning period for route 39 and the Evening period for route 42 are not expected, it is expected that the two peak periods and the Base period perform reasonably well because of the large sample size in those periods. Overall, other than the Early Morning period, the results of route 42 are quite encouraging. However, the absolute percentage difference between the ride check information and the electronic estimates is still quite large: it varies between 1 and 122%.

From the above discussion, it is clear that this method works quite poorly with the small routes during the Early Morning and the Evening periods. However, the method works reasonably well with the large routes during the AM Peak, the Base, and the PM

Peak Periods, and the results of 1996 are considerably better than those of 1994. Therefore, with continued driver compliance improvement, we can see that the method has the potential to be used to estimate ridership by route and time period for the large routes and even some medium-sized routes for the AM Peak, Base, and PM Peak periods. The results of the AM Peak period are the most encouraging. On the other hand, it is clear that the method cannot be used for the Early Morning and Evening periods at this time because there are too few trips in these two periods, and the ridership fluctuates quite a bit. However, with continued improvement in driver compliance, it is not totally impossible for the method to satisfactorily estimate ridership by route and time period for these two periods for the relatively large routes.

Chapter 5

Conclusions

This chapter presents the conclusions of this thesis and describes some possible further research directions.

5.1 Summary and Conclusions

This thesis contains two major components:

1. A survey of the experience of transit agencies in the United States with electronic fareboxes
2. Develop and test a number of methods to estimate various revenue and ridership information at different levels of detail from electronic farebox data gathered at the MBTA

The following subsections describe these two components in turn.

5.1.1 The Experience of Transit Agencies with Electronic Fareboxes

A survey was conducted of 8 (plus the MBTA) medium-sized and large transit agencies in the United States to ascertain their experience with electronic fareboxes. The agencies surveyed ranged from those having considerable experience with the electronic farebox to those just beginning to use electronic farebox data to estimate revenue and ridership information. From the surveys, several conclusions can be drawn.

Most agencies agreed that the level of driver compliance is directly related to success in using the electronic farebox for planning and service evaluation purposes. Transit agencies successful with the electronic farebox are generally those who either have considerable experience with it or have a good and comprehensive training, retraining, and disciplinary program for operators. Most transit agencies use revenue data for planning or service evaluation purposes, but few uses ridership data for the same purposes, and these are usually limited to agencies which are more successful with the electronic farebox. However, none of the surveyed agencies currently uses electronic fareboxes to estimate maximum passenger load to any extent. Most agencies use an adjustment factor to adjust the raw revenue and ridership figures recorded by the electronic farebox, and the garage revenue count is generally used as the control total for estimating revenue information whereas ride checks are used to validate electronic farebox ridership information. Some agencies are trying to use the electronic farebox to complement ride checks. Generally, these agencies would first look at electronic farebox ridership data to see if there has been any significant change in ridership, and if there is, ride checks would be performed to confirm the change before any action is taken. Some transit agencies have unfulfilled goals. These goals include the integration of the bus and the subway fare systems, having better and faster access to different levels of detail of revenue and ridership information, and improvement in driver compliance with the proper keypad entry procedures.

5.1.2 The Proposed Methods for Estimating Revenue and Ridership at Different Levels of Detail and the Test Results

This thesis proposed and tested a number of methods to estimate revenue and ridership at different levels of detail. Basically, four types of information were analyzed:

1. Daily Route Fare Revenue
2. Daily Route Pass Swipes by Type (Local Bus and Combo Passes)
3. Daily Route Ridership
4. Ridership by Route and Time Period

5.1.2.1 Proposed Methods

For the estimation of daily route fare revenue, four methods are proposed. Method 1 allocates the unknown fare revenue to each route in proportion to their recorded fare revenue. Method 2 allocates the unknown fare revenue to each route based on missing trip records. Method 3 screens out bad trip records and then applies Method 2. Method 4 combines Methods 1 and 3.

For the estimation of daily route pass swipes by type (local bus and combo passes), only three methods were proposed, closely paralleling the first three methods proposed for estimating daily route revenue. Method 4 is not proposed for estimating daily route pass swipes by type because Method 3 performs so well that Method 1 is not needed, even for small routes, making Method 4 identical to Method 3.

For the estimation of daily route ridership, a method which does not depend on a control total for ridership is proposed. This method depends on the definition of average fare and total pass swipes and uses Method 3 for estimating daily route fare revenue.

For the estimation of route ridership by time period, a two-part method is proposed. The first part involves estimating the daily route ridership by applying the previous method. The second part allocates this ridership across the five periods of the day with the daily ridership serving as the control total. This part also attempts to handle the errors caused by unregistered fare riders and missing trip records.

The major difference between Method 3 and the first two methods is that Method 3 introduces the concept of trip-level screening. This trip-level screening attempts to screen out trip records with one or more of the following three criteria:

1. Having a defective electronic farebox
2. Spaced too close to the following trip record
3. Spaced too far from the following trip record

A defective farebox is a farebox which fails to record fare revenue or pass swipes. Typically this affects only a very small percentage (less than 2%) of all scheduled trips. However, this screening is necessary because some small routes have only one or two buses running the route, and if a bus has a defective electronic farebox, the quality of the data for this small route will be severely affected.

The second and third criteria depend on the concept of inter-record times. Inter-record time is defined as the amount of time between the creation of two consecutive farebox records. If two records are very close to each other (short inter-record time), it is likely that both records contain partial information of the same trip due to multiple registrations by the operator. In this case, the data in these two records are combined, and the record time is set to the earlier of the two record times. This problem occurs more often than the defective farebox problem. Generally 5% to 7% of the number of scheduled trips have this problem. If two records are spaced too far from each other (long inter-record time), it is likely that the earlier record contains information for more than one trip due to missing registrations. In this case, the earlier record is screened out. This is the most common problem among the three mentioned. In 1994, 30% to 48% of the number of scheduled trips have this problem, and in 1996, with improved driver compliance, 19% to 21% of the number of scheduled trips have this problem.

5.1.2.2 Test Results

The tests are very straightforward. Data provided by the Massachusetts Bay Transportation Authority on Bartlett Garage are used to test the proposed methods. Methods 1 and 2 use a sample size of 40 whereas Method 3 (and 4) uses a sample size of 5 due to their complexity. Since revenue and ridership are not expected to vary substantially across seasons, which is confirmed by analyzing the control totals, the estimated item for the winter is compared with that of the spring to determine if they are statistically different. If they are not, the method is considered successful in analyzing the route; otherwise, it is considered unsuccessful. Both 1994 and 1996 data are analyzed to study the effects of different levels of driver compliance. In addition, the coefficients of variation are also studied to determine the stability of data within a season. Generally a coefficient of variation greater than 0.3 is considered high. However, since the sample size is very small for Methods 3 and 4 (5), it is likely that the coefficients of variation will change as the sample size increases. The pass swipe analysis is performed slightly differently. Since the control totals are statistically different in many cases, most pass swipe analyses are performed by comparing estimated pass swipe shares instead of the estimated pass swipes since pass swipe shares are not expected to vary significantly across seasons.

In estimating daily route fare revenue, Method 4 clearly performs better than the other three methods. First, it generally produces lower coefficients of variation than the other methods, but it does not necessarily produces higher t-values than the z-values produced by Methods 1 and 2. In fact, fewer routes have mean estimated fare revenue considered statistically different between the winter and the spring when analyzed by Method 4 instead of by Methods 1 or 2. Second, Method 4 works better than Methods 1 and 2 on medium-sized routes. Finally, Method 4, unlike Method 3, allocates some fare revenue by means of Method 1 to routes which are left with no data after the screenings. This is important not only to these routes, but to other routes as well. Since Method 3 does not allocate any fare revenue to the routes with no data after the screenings, all of these fare revenue is allocated to the other routes. However, since some of these fare revenue actually belong to those routes which are left with no data after the screenings, the estimated fare revenue of other routes is probably inflated.

In estimating daily route pass swipes, Method 3 clearly outperforms the other two methods for both local bus and combo pass analyses. Similar to the fare revenue analysis, Method 3 produces lower coefficients of variation and fewer routes with mean estimated pass swipe shares statistically different between the winter and the spring.

In estimating daily route ridership, only one method is proposed. Although this method generally produces low coefficients of variation, quite a few routes have statistically different mean estimated daily route ridership between the winter and the spring, and the difference between the ride check ridership information and the electronic farebox ridership estimate for route 39 is quite large. However, the results for route 42 are quite encouraging. Since the sample size used in this analysis is small, and only two routes have ride check data available for comparison, it is still not very clear if this method is capable of estimating daily ridership by route.

In estimating ridership by route and time period, again, only one method is proposed. This method performs quite poorly with the small routes during the Early Morning and the Evening periods. The coefficients of variation are generally very high, and many small routes have all data eliminated after the screening, especially for 1994 and winter 1996. However, the spring 1996 results are quite encouraging, and those for the AM Peak, the Base, and the PM Peak periods for the large routes are quite good. Also, the ride check comparison for route 42 is very satisfactory. Overall, among the five time periods, the AM Peak, Base, and PM Peak periods have better estimates than the other two periods.

5.1.3 Conclusions

Overall, a number of conclusions can be drawn. First, almost all methods work better with large routes than with small routes because the large routes generally have more data. With more data, it is easier to find enough good data, which is essential for the methods to work. Although with more data there will be more bad data, they are likely to be eliminated by the screening procedure of Method 3 or 4. This shows that having a large sample size is very important to the success of the methods. In fact, if the sample size used for the analyses of Methods 3 and 4 were increased, the results might have been much better. Second, since in all cases Method 3 or 4 work better than Methods 1 and 2, it is clear that trip-level screening is essential to the success of the estimation. Finally, driver compliance is very important. At the MBTA, driver compliance has improved significantly in 1996. In almost all cases, the 1996 data work significantly better than the 1994 data. In fact, in the pass swipe analysis, Method 4 is no longer necessary because the driver compliance is so good that no route, even the small ones, has no data left after the screening procedure of Method 3. In addition, the daily probing of the farebox is essential too because this produces reliable control totals, which are important for all the proposed methods. Finally, this research shows that using the electronic farebox data to estimate daily fare revenue and pass swipes by route is more plausible than to estimate route ridership by day and time period at this stage.

5.2 Future Work

Given the findings of this research, we can see that additional research in some areas may help us understand how to use electronic fareboxes to better estimate revenue and ridership information for planning and service evaluation purposes.

First, although this research devotes a significant amount of effort to trip-level screening, it is not the only focus of this research. In order to keep the amount of work manageable, a small sample size (5) is used in all tests involving trip-level screening. Unfortunately, such a small sample size does limit the amount of information we can get from the results. For example, we cannot put much confidence on the coefficients of variation because of the small sample size, which makes our 0.3 cutoff somewhat suspect. Since we now see that trip-level screening is essential to the success of the estimation, it would be extremely useful to test Methods 3 and 4 for all kinds of information with a larger sample size, such as 40. This would make our 0.3 cutoff more defensible, and probably would result in fewer routes with no data remaining after the screening procedure.

Second, we see that driver compliance is extremely important to the success of the methods. Although the tests with 1996 data try to incorporate the effects of improved driver compliance, the significant driver compliance improvement occurred in the spring of 1996, which means that the winter 1996 data do not incorporate the effects of improved driver compliance. Therefore, it would be useful to test the methods completely with data from Spring 1996 and after. For example, a test using Spring 1996 and Fall 1996 data would be interesting.

Third, in this study the median inter-record times, unlike the average fares which are based only on spring 1996 data, are calculated using data from the same season. However, if the whole season is bad (this is likely for the small routes given our small sample size), it is quite likely that the calculated median inter-record times are unreliable. Therefore, it would be useful to analyze data from several seasons to establish a set of universal median inter-record times for all routes which are independent of the season. This would produce a set of less biased median inter-record times. In fact, this strategy should also be applied to average fares because even though the spring 1996 estimated average fares are good, future ones may be better.

Fourth, since we are conservative in this study, the screening procedure routinely eliminates the last trip of a bus because there is no inter-record time available for these trips. Although this does not represent a large number of trips, many good trip records are certainly eliminated for this reason. It would be useful to develop computer programs which can keep track of all the good drivers so as not to eliminate the last trip for these good drivers. This would save us some good trip records.

Finally, this thesis analyzes only the routes of Bartlett Garage on weekdays. This is based on the assumption that if the methods work well with the routes in a large garage like Bartlett, they would probably work for all routes. Also, this is to reduce unnecessary complexity. In order to confirm this assumption, it would be useful to test the methods on routes from another garage and to try some weekend data.

Appendix A

MBTA Electronic Farebox Reports

Table A.1 shows a sample MBTA Revenue and Ridership Summary by Route. The first column shows the signcodes. Although the summary itself names the column as route number, the column actually contains the signcodes, not the route numbers. The signcode specifies the route number and the variation of the trip. For example, the signcode 240 means the variation 0 of route 24, and the signcode 336 means the variation 6 of route 33. However, there are two exceptions. First, although the signcode 39 is an invalid signcode, the operators type it in so frequently that the system is programmed to recognize this signcode as a variation of route 39. Unfortunately, this signcode does not specify the variation of the route. The other exception is that the signcodes 342 through 349 specify the variations of route 34E instead of those of route 34 most of the time. The reason for having two routes sharing one signcode is that there are more than 9 variations for routes 34 and 34E altogether. Since the signcodes are originally designed for the destination signs (not the electronic fareboxes), when some of the trips of two routes terminate at a common terminal via a common point, one (instead of two) signcode can be used. The second columns shows the number of trips recorded for each signcode. The third column shows the recorded fare revenue (route revenue) for each signcode for the day of interest. This recorded fare revenue includes the unclassified revenue as well. Unclassified revenue is the revenue collected in the farebox that could be allocated to a trip, but cannot be allocated to any particular fare category of that trip. One possible source of unclassified revenue is the money that was dumped into the farebox vault as a result of the operator pressing the green dump button. The last item in this column is the total recorded fare revenue for the garage including all unknown and "others" fare revenue. This item is also called the fare revenue control total for the garage for the day. The fourth through eighth columns show the total recorded ridership, the total number of tokens collected, the total number of tickets collected, the total number of pass swipes, and the total number of envelopes received for each signcode for the day, respectively. The last two columns contain the unclassified revenue and the percentage of route revenue that is unclassified revenue, respectively. In the analysis of fare revenue using Method 1, only the first (signcode) column and the third (route revenue) column are used. Method 2 uses the second (number of trip records) column in additional to the first and the third columns.

Table A.2 shows a sample Trip-Level Revenue and Ridership Summary. The first column contains the signcode of each trip. Again, although the summary itself names the column as route number, the column actually contains the signcode of each trip, not the route number. The second and third columns contain the run and bus numbers of each trip, respectively. Please note that the bus number is not entered by the operator but automatically generated by each bus; hence it is not affected by driver entry errors. The fourth column contains the date and time of each trip. The fifth and sixth columns contain the current revenue (recorded fare revenue) and the unclassified revenue of each trip, respectively. Please note that the current revenue (recorded fare revenue) of each trip includes the unclassified revenue of that trip. Therefore, the unclassified revenue of each trip can never exceed the current revenue of that trip. The seventh column contains the number of times the operator pressed the dump button during each trip. The eighth column contains the fareset number of each trip, which is the code for the regular adult fare of that route. For example, if the regular adult fare of route 28 (signcode 280 or 281) is 60 cents, then the fareset number (the code for the regular adult fare) is 1. The ninth column contains the fare count of each trip, which is the number of regular adult fares collected

Table A.1: A sample MBTA Revenue and Ridership Summary by Route

DAILY ROUTE SUMMARY FOR 02-JAN-1996

MBTA Operations

03 JAN 1996 | 02:37 PM | PAGE 2

FOR: Bartlett

Totals are listed for each valid route number. Values for routes not found on the list of valid numbers are included in OTHER.

The farebox normally keeps revenue and ridership values on a route-level basis, but if it uses up its maximum of 100 route/run records, it can no longer do so. Any remaining totals are included under the UNKNOWN route number.

REVENUE AND RIDERSHIP SUMMARY BY ROUTE

ROUTE NUMBER	NUMBER OF RUNS	ROUTE REVENUE	TOTAL RIDERSHIP	TOTAL TOKENS	TOTAL TICKETS	TOTAL PASSES	TOTAL ENVELOPES	UNCLASSIFIED REVENUE	UNCLASSIFIED vs CURRENT REVENUE
39	10	58.91	287	3	0	176	0	4.76	8.08 %
140	5	28.25	103	2	0	33	0	5.06	17.91 %
141	4	10.34	32	0	0	11	0	1.59	15.38 %
210	3	2.71	10	0	0	5	0	2.01	74.17 %
240	19	132.58	397	0	0	160	0	20.90	15.76 %
241	18	33.76	162	0	0	33	0	1.38	4.09 %
242	1	5.53	12	0	0	1	0	0.05	0.90 %
260	12	41.56	121	2	0	36	0	7.88	18.96 %
261	6	117.46	359	2	0	144	3	9.33	7.94 %
270	10	19.37	61	0	0	26	0	2.72	14.04 %
271	8	31.76	95	0	0	41	0	1.31	4.12 %
280	104	740.04	2332	11	0	832	3	98.09	13.25 %
281	107	781.16	2632	7	0	989	6	92.87	11.89 %
282	2	2.96	9	0	0	5	0	0.71	23.99 %
290	31	150.19	467	2	0	200	1	21.19	14.11 %
291	19	66.46	236	1	0	77	1	8.22	12.37 %
295	18	153.33	480	0	0	173	0	24.69	16.10 %
300	6	90.77	231	3	0	81	2	38.62	42.65 %
301	9	77.30	308	3	0	101	0	7.16	9.26 %
302	1	2.40	7	0	0	0	0	1.80	75.00 %
303	1	0.00	1	0	0	1	0	0.00	0.00 %
310	68	465.45	1538	3	0	691	1	50.46	10.84 %
311	60	293.67	1121	6	0	483	0	40.52	13.80 %
320	87	671.24	2260	6	0	818	0	53.27	7.94 %
321	76	549.15	1861	6	0	769	2	38.66	7.04 %
322	24	149.36	584	3	0	204	1	11.19	7.49 %
330	1	21.98	69	0	0	31	0	1.98	9.01 %
332	10	9.05	25	0	0	13	0	2.90	32.04 %
334	1	3.30	10	0	0	4	0	0.00	0.00 %
336	8	19.30	56	1	0	19	0	2.95	15.13 %
340	49	488.46	1582	5	0	595	0	41.91	8.58 %
341	42	291.72	908	6	0	336	1	32.42	11.11 %
343	5	52.49	149	1	0	40	1	3.89	6.90 %

DAILY ROUTE SUMMARY FOR 03 JAN 1996
MBTA Operations

11.JAN 1996 | 02:38 PM | PAGE 2

FOR: Bartlett

REVENUE AND RIDERSHIP SUMMARY BY ROUTE

ROUTE NUMBER	NUMBER OF BUS	ROUTE REVENUE	TOTAL RIDERSHIP	TOTAL TOKENS	TOTAL TICKETS	TOTAL PASSES	TOTAL ENVELOPES	UNCLASSIFIED REVENUE	UNCLASSIFIED vs CURRENT REVENUE
[illegible]	24	[illegible]	[illegible]	0	0	[illegible]	0	9.83	8.17 %
[illegible]	9	[illegible]	[illegible]	0	0	[illegible]	0	11.83	14.14 %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]	2.50	9.40 %
[illegible]	[illegible]	[illegible]	[illegible]	3	0	[illegible]	[illegible]	23.24	7.69 %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	42.86	[illegible] %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]	2.60	3.74 %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]	19.44	5.27 %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]	4.06	6.22 %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	25	3.10	[illegible] %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	17.95	[illegible] %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.78	[illegible] %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible] %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.77	[illegible] %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	21.39	[illegible] %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	197	[illegible]	13.74	9.94 %
230	22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	11.90	7.51 %
231	[illegible]	41.76	[illegible]	0	[illegible]	[illegible]	0	3.06	7.33 %
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	130.62	[illegible] %
[illegible]	[illegible]	[illegible]	[illegible]	41	[illegible]	[illegible]	[illegible]	126.40	7.23 %
232	[illegible]	21.09	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.01	8.34 %
237	[illegible]	9.35	[illegible]	0	[illegible]	[illegible]	0	0.65	6.60 %
[illegible]	13	53.49	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	5.78	9.72 %
401	16	71.76	[illegible]	0	0	31	1	6.73	9.38 %
410	11	47.28	[illegible]	1	0	56	0	2.50	5.29 %
411	3	10.59	22	0	0	9	0	0.05	0.47 %
420	32	[illegible]	[illegible]	3	[illegible]	[illegible]	[illegible]	48.20	15.02 %
421	25	[illegible]	[illegible]	4	[illegible]	[illegible]	[illegible]	32.45	12.93 %
422	2	6.10	[illegible]	[illegible]	0	[illegible]	0	1.90	31.15 %
423	1	[illegible]	7	0	[illegible]	[illegible]	0	0.00	0.00 %
460	24	65.27	[illegible]	0	0	[illegible]	[illegible]	6.67	10.22 %
461	[illegible]	[illegible]	[illegible]	3	0	22	0	2.66	8.68 %
480	[illegible]	34.46	[illegible]	[illegible]	0	[illegible]	[illegible]	4.15	12.04 %
500	15	72.37	[illegible]	2	0	102	0	5.97	8.25 %
501	11	17.43	[illegible]	0	[illegible]	[illegible]	0	0.80	4.57 %
510	24	[illegible]	610	2	0	[illegible]	0	24.30	12.54 %
511	16	142.24	447	[illegible]	0	[illegible]	10	7.63	5.34 %
520	7	61.86	[illegible]	1	0	125	0	4.46	7.21 %
521	5	10.40	[illegible]	[illegible]	0	18	0	1.35	12.98 %
522	5	15.25	42	0	0	16	0	1.01	6.62 %
523	1	1.20	4	0	0	[illegible]	0	0.00	0.00 %
524	5	5.00	11	0	0	1	0	0.50	10.00 %
525	4	11.25	[illegible]	[illegible]	0	19	0	1.80	16.00 %
529	1	9.95	25	0	0	8	0	0.20	2.01 %
1607	1	68.27	147	1	0	61	0	17.87	26.18 %

DAILY ROUTE SUMMARY FOR 03 JAN 1996
MBTA Operations

11 JAN 1996 02:38 PM PAGE 3

FOR: Bartlett

REVENUE AND RIDERSHIP SUMMARY BY ROUTE

ROUTE NUMBER	NUMBER OF RUNS	ROUTE REVENUE	TOTAL RIDERSHIP	TOTAL TOKENS	TOTAL TICKETS	TOTAL PASSES	TOTAL ENVELOPES	UNCLASSIFIED REVENUE	UNCLASSIFIED vs CURRENT REVENUE
[illegible]910	[illegible]	[illegible]	[illegible]	0	0	[illegible]	[illegible]	2.11	21.62 %
OTHERS	[illegible]	[illegible]	[illegible]	13	0	[illegible]	30	224.78	16.67 %
UNKNOWN	....	[illegible]	[illegible]	12	0	[illegible]	[illegible]	107.13	3.23 %
TOTALS	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1319.02	12.52 %

TOKENS, TICKETS AND PASSES BY ROUTE

[illegible]	TYP 1	TYP 2	TYP 3	TYP 4	TYP 5	TYP 6	TYP 7	TYP 8	TYP 9	TYP 10	TYP 11	TYP 12
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
141	0	[illegible]	[illegible]	0	[illegible]	0	[illegible]	[illegible]	[illegible]	0	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
240	0	0	32	0	68	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]
241	[illegible]	[illegible]	[illegible]	[illegible]	3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
242	[illegible]	0	[illegible]	[illegible]	0	0	0	[illegible]	[illegible]	0	[illegible]	[illegible]
250	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	[illegible]
251	2	[illegible]	[illegible]	[illegible]	96	2	0	[illegible]	[illegible]	0	0	[illegible]
270	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2	[illegible]	0	[illegible]	0	[illegible]	[illegible]
271	0	0	[illegible]	0	[illegible]	1	0	0	0	0	0	0
280	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]	[illegible]	[illegible]	0	0	[illegible]	[illegible]
281	7	0	682	[illegible]	290	12	[illegible]	2	[illegible]	0	0	[illegible]
282	[illegible]	0	[illegible]	[illegible]	3	0	[illegible]	0	[illegible]	[illegible]	0	[illegible]
290	2	[illegible]	[illegible]	[illegible]	[illegible]	3	[illegible]	[illegible]	0	0	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]	[illegible]	[illegible]	0
295	0	[illegible]	97	0	[illegible]	[illegible]	[illegible]	0	0	[illegible]	[illegible]	[illegible]
300	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2	[illegible]	0	0	[illegible]	0	[illegible]
301	[illegible]	0	[illegible]	[illegible]	44	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0
302	[illegible]	[illegible]	0	[illegible]	0	0	[illegible]	0	0	0	[illegible]	0
303	0	0	[illegible]	0	0	0	0	0	0	[illegible]	0	0
310	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]
311	[illegible]	0	[illegible]	0	[illegible]	4	2	[illegible]	[illegible]	0	0	[illegible]
320	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
321	6	[illegible]	[illegible]	[illegible]	[illegible]	13	22	[illegible]	2	0	0	0
322	[illegible]	0	[illegible]	0	[illegible]	5	[illegible]	[illegible]	[illegible]	[illegible]	0	0
330	0	0	12	0	13	2	4	0	0	0	0	0

Table A.2: A sample Trip-Level Revenue and Ridership Summary

SUMMARY FOR ROUTE 280 FOR 26 JUN 1996
MBTA Operations
01 JUL 1996 06:00 PM PAGE 1

FOR: Bartlett

Report Sequence : by trip date and time within route
Selected Date(s): 26 JUN 1996
Selected Input : Farebox data for 26 - 30 JUN 1996
Report includes trip records having no revenue or riders.
The following data files were not available:
27 JUN 1996 for Bartlett
29 JUN 1996 for Bartlett
30 JUN 1996 for Bartlett

REVENUE AND RIDERSHIP SUMMARY

ROUTE #	RUN #	BUS #	DATE & TIME	CURRENT REVENUE	UNCLSFD REVENUE	DUMP COUNT	FARE SET#	FARE COUNT	TOTAL TOKENS	TOTAL TICKETS	TOTAL PASSES	EMLP COUNT	TOTAL RIDERS
280	1139	214	26 JUN 00:00	0.00	0.00	0	1	0	0	0	0	0	0
280	1149	252	26 JUN 00:00	0.25	0.00	0	1	0	0	0	1	0	3
280	1140	202	26 JUN 00:02	0.93	0.93	4	1	0	0	0	7	0	7
280	1149	252	26 JUN 01:06	1.10	0.00	0	1	3	0	0	1	0	9
280	1026	240	26 JUN 04:55	1.55	0.00	0	1	1	0	0	6	0	10
280	1019	208	26 JUN 05:39	1.05	0.00	0	1	4	0	0	6	0	12
280	1010	228	26 JUN 05:44	1.30	0.00	0	1	0	0	0	7	0	10
280	1020	216	26 JUN 05:57	1.77	0.32	4	1	3	0	0	7	0	23
280	1010	228	26 JUN 05:58	1.00	0.00	0	1	0	0	0	0	0	0
280	1026	240	26 JUN 06:13	1.83	0.00	0	1	4	0	0	17	0	30
280	1023	8667	26 JUN 06:16	2.10	0.00	0	1	3	0	0	5	0	14
280	1072	218	26 JUN 06:29	0.00	0.00	0	1	0	0	0	0	0	0
280	1010	228	26 JUN 06:33	4.40	1.95	7	1	3	0	0	12	0	29
280	1050	8681	26 JUN 06:40	5.39	0.29	2	1	7	0	0	11	0	30
280	1019	208	26 JUN 06:49	4.11	0.05	1	1	4	0	0	8	0	41
280	1058	229	26 JUN 06:57	2.55	0.00	0	1	2	0	0	12	0	33
280	1073	195	26 JUN 07:04	5.18	0.48	2	1	5	0	0	5	0	18
280	1020	216	26 JUN 07:06	6.15	0.25	3	1	6	0	0	2	0	39
280	1010	228	26 JUN 07:11	0.00	0.00	0	1	0	0	0	0	0	0
280	1010	228	26 JUN 07:11	0.00	0.00	0	1	0	0	0	0	0	0
280	1026	240	26 JUN 07:25	1.75	0.05	1	1	2	0	0	7	0	19
280	1069	8686	26 JUN 07:39	7.42	0.02	1	1	7	0	0	10	0	32
280	1072	218	26 JUN 07:45	10.61	0.22	6	1	12	0	0	12	0	43
280	1023	8667	26 JUN 07:45	2.80	0.25	1	1	2	0	0	3	0	11
280	1010	228	26 JUN 07:57	7.51	2.71	10	1	8	0	0	17	0	29
280	1050	8681	26 JUN 08:05	2.30	0.80	4	1	2	0	0	4	0	9
280	1058	229	26 JUN 08:16	7.01	0.65	2	1	7	0	0	18	0	34
280	1019	208	26 JUN 08:17	5.55	0.25	1	1	4	0	0	12	0	29
280	1073	195	26 JUN 08:25	4.66	0.51	2	1	4	0	0	12	0	22
280	1020	216	26 JUN 08:30	6.16	0.30	3	1	6	0	0	14	0	49
280	1010	228	26 JUN 08:40	0.00	0.00	0	1	0	0	0	0	0	0
280	1026	240	26 JUN 08:41	6.10	0.00	0	1	5	0	0	7	0	30
280	1069	8686	26 JUN 08:57	5.28	0.05	1	1	5	0	0	11	0	33

Table A.3: A sample MBTA Tokens, Tickets, and Passes by Route Report

TOKENS, TICKETS AND PASSES BY ROUTE

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	TYP 11	TYP 12
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	32	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	36	2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0
[illegible]	[illegible]	[illegible]	37	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2	[illegible]	[illegible]	[illegible]	[illegible]	0	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	0	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	0	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]	0	[illegible]	[illegible]	[illegible]	[illegible]	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	13	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]

DAILY ROUTE SUMMARY FOR 03 JAN 1996
MBTA Operations

11 JAN 1996 02:38 PM PAGE 4

FOR: Bartlett

TOKENS, TICKETS AND PASSES BY ROUTE

ROUTE	TYP 1	TYP 2	TYP 3	TYP 4	TYP 5	TYP 6	TYP 7	TYP 8	TYP 9	TYP 10	TYP 11	TYP 12
[illegible]	[illegible]	[illegible]	5	[illegible]	5	0	[illegible]	[illegible]	0	[illegible]	0	[illegible]
[illegible]	[illegible]	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	0
344	[illegible]	[illegible]	[illegible]	[illegible]	46	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0
345	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
346	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
347	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	35	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	65	0	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	0	1193	[illegible]	471	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	0
[illegible]	41	[illegible]	1221	[illegible]	396	15	[illegible]	[illegible]	50	[illegible]	0	0
[illegible]	[illegible]	0	[illegible]	0	[illegible]	0	0	[illegible]	[illegible]	0	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	11	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]
400	[illegible]	0	[illegible]	0	[illegible]	0	[illegible]	[illegible]	[illegible]	0	0	0
[illegible]	[illegible]	0	[illegible]	[illegible]	43	0	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]
410	[illegible]	0	43	[illegible]	23	0	0	0	[illegible]	[illegible]	0	[illegible]
411	0	[illegible]	7	0	2	0	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
420	[illegible]	[illegible]	[illegible]	[illegible]	38	10	0	[illegible]	[illegible]	0	0	0
421	4	0	[illegible]	[illegible]	30	3	[illegible]	4	[illegible]	[illegible]	0	[illegible]
422	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	0	0	0	0	0
423	0	0	[illegible]	[illegible]	[illegible]	0	0	[illegible]	[illegible]	0	0	[illegible]
460	0	[illegible]	[illegible]	0	16	2	0	[illegible]	[illegible]	0	0	[illegible]
461	3	0	[illegible]	0	[illegible]	0	0	[illegible]	[illegible]	0	[illegible]	[illegible]
480	1	0	12	0	5	0	0	0	0	0	0	0
500	2	0	38	0	51	1	10	[illegible]	0	0	0	0
501	0	0	[illegible]	0	13	0	0	1	[illegible]	0	0	0
510	2	0	124	0	134	2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
511	[illegible]	0	63	0	81	3	9	0	0	0	0	[illegible]
520	[illegible]	0	38	0	76	6	4	0	[illegible]	0	0	[illegible]
521	0	0	[illegible]	0	10	0	[illegible]	0	0	0	0	0
522	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

DAILY ROUTE SUMMARY FOR 03 JAN 1996
MBTA Operations
11 JAN 1996 02:38 PM PAGE 5

FOR: Bartlett

TOKENS, TICKETS AND PASSES BY ROUTE

ROUTE	TTP 1	TTP 2	TTP 3	TTP 4	TTP 5	TTP 6	TTP 7	TTP 8	TTP 9	TTP 10	TTP 11	TTP 12
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	0
[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	0	[illegible]	[illegible]	[illegible]	0	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
OTHERS	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]
[illegible]	[illegible]	[illegible]	185	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0
TOTALS	227	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0

during each trip. Finally, the tenth through the fourteenth columns contain the total number of tokens, tickets, passes, envelopes, and riders for each trip, respectively. In the analysis of fare revenue using Method 3, only the first five columns are needed.

Table A.3 shows a sample MBTA Tokens, Tickets, and Passes by Route Report. The first column shows the signcodes. The remaining columns show the different kinds of tokens, tickets, and passes registered. We will focus on two items: local bus pass swipes and combo pass swipes, which are contained in the fourth (TTP 3) and sixth (TTP 5) columns, respectively. The last item in the fourth column is the total recorded local bus pass swipes for the garage on the day of interest. This item is also called the control total for the local bus pass for the garage for that day. Similarly, the last items on the sixth column contain the control totals for the combo pass for the garage for the day of interest.

Table A.4 shows a sample Trip-Level Tokens, Tickets, and Passes Report. The first column contains the signcode of each trip. The second and third columns contain the run and bus numbers of each trip, respectively. The fourth column contains the date and time of each trip. The remaining twelve columns contain the number of different kinds of tokens, tickets, and passes registered. Again, only local bus pass swipes and combo pass swipes are analyzed, and their swipes are recorded on columns 7 (TTP 3) and 9 (TTP 5), respectively.

Table A.4: A sample Trip-Level Tokens, Tickets, and Passes Report

SUMMARY FOR ROUTE 280 FOR 26 JUN 1996
MBTA Operations

01 JUL 1996 06:00 PM PAGE 7

FOR: Bartlett

TOKENS, TICKETS AND PASSES

ROUTE #	RUN #	BUS #	DATE & TIME	TTP 1	TTP 2	TTP 3	TTP 4	TTP 5	TTP 6	TTP 7	TTP 8	TTP 9	TTP 10	TTP 11	TTP 12
280	1020	216	26 JUN 05:57	0	0	6	0	1	0	0	0	0	0	0	0
280	1010	228	26 JUN 05:58	0	0	0	0	0	0	0	0	0	0	0	0
280	1026	240	26 JUN 06:13	0	0	15	0	2	0	0	0	0	0	0	0
280	9023	8667	26 JUN 06:16	0	0	3	0	1	0	0	1	0	0	0	0
280	1072	218	26 JUN 06:29	0	0	0	0	0	0	0	0	0	0	0	0
280	1010	228	26 JUN 06:33	0	0	10	0	2	0	0	0	0	0	0	0
280	1050	8681	26 JUN 06:40	0	0	8	0	3	0	0	0	0	0	0	0
280	1019	208	26 JUN 06:49	0	0	6	0	1	1	0	0	0	0	0	0
280	1058	229	26 JUN 06:57	0	0	10	0	2	0	0	0	0	0	0	0
280	1073	195	26 JUN 07:04	0	0	5	0	0	0	0	0	0	0	0	0
280	1020	216	26 JUN 07:06	0	0	2	0	0	0	0	0	0	0	0	0
280	1010	228	26 JUN 07:11	0	0	0	0	0	0	0	0	0	0	0	0
280	1010	228	26 JUN 07:11	0	0	0	0	0	0	0	0	0	0	0	0
280	1026	240	26 JUN 07:25	0	0	5	0	2	0	0	0	0	0	0	0
280	1069	8686	26 JUN 07:39	0	0	5	0	4	0	0	0	0	0	0	0
280	1072	218	26 JUN 07:45	0	0	9	0	3	0	0	0	0	0	0	0
280	9023	8667	26 JUN 07:45	0	0	3	0	0	0	0	0	0	0	0	0
280	1010	228	26 JUN 07:57	0	0	11	0	6	0	0	0	0	0	0	0
280	1050	8681	26 JUN 08:05	0	0	3	0	1	0	0	0	0	0	0	0
280	1058	229	26 JUN 08:16	0	0	15	0	3	0	0	0	0	0	0	0
280	1019	208	26 JUN 08:17	0	0	9	0	2	1	0	0	0	0	0	0
280	1073	195	26 JUN 08:25	0	0	6	0	6	0	0	0	0	0	0	0
280	1020	216	26 JUN 08:30	0	0	9	0	4	0	0	0	1	0	0	0
280	1010	228	26 JUN 08:40	0	0	0	0	0	0	0	0	0	0	0	0
280	1026	240	26 JUN 08:41	0	0	4	0	2	1	0	0	0	0	0	0
280	1069	8686	26 JUN 08:57	0	0	9	0	2	0	0	0	0	0	0	0
280	1072	218	26 JUN 09:04	0	0	1	0	4	1	0	0	0	0	0	0
280	9023	8667	26 JUN 09:06	0	0	14	0	2	0	0	1	1	0	0	0
280	1078	251	26 JUN 09:07	0	0	6	0	3	0	0	0	1	0	0	0
280	1050	8681	26 JUN 09:22	0	0	7	0	2	1	0	0	0	0	0	0
280	1019	208	26 JUN 09:30	0	0	21	0	10	1	0	0	0	0	0	0
280	1073	195	26 JUN 09:43	0	0	23	0	7	0	0	0	0	0	0	0
280	1023	214	26 JUN 10:04	0	0	13	0	1	0	0	0	0	0	0	0
280	1089	8757	26 JUN 10:10	0	0	12	0	5	0	0	0	0	0	0	0
280	1072	218	26 JUN 10:27	0	0	10	0	2	1	0	0	0	0	0	0
280	1078	251	26 JUN 10:36	0	0	10	0	5	0	0	0	0	0	0	0
280	1007	229	26 JUN 11:05	0	0	16	0	6	0	0	1	0	0	0	0
280	1073	195	26 JUN 11:23	0	0	14	0	3	0	0	0	0	0	0	0
280	1089	8757	26 JUN 11:27	0	0	14	0	5	0	0	0	0	0	0	0
280	1107	214	26 JUN 11:50	0	0	18	0	3	1	0	0	1	0	0	0
280	1114	218	26 JUN 11:52	0	0	10	0	2	0	0	0	0	0	0	0
280	1078	251	26 JUN 11:59	0	0	19	0	3	0	0	0	0	0	0	0
280	1031	241	26 JUN 12:09	0	0	[illegible]	0	7	0	0	0	0	0	0	0

Appendix B

The MBTA Electronic Farebox Quick Reference Guide

Figure B.1 shows the MBTA Electronic Farebox Quick Reference Guide. In order to remind drivers of the proper farebox entry procedure, this quick reference guide was given to each driver, with different versions available for buses, trackless trolleys, and streetcars. It lists all keypad labels and tells drivers how to classify the different types of passengers and press the right keys on the electronic farebox. It also reminds the drivers to enter the signcode after each half roundtrip, reviews the proper log on, signcode, and log out procedures, and discusses the fare exceptions such as transfers.

Figure B.1: The MBTA Electronic Farebox Quick Reference Guide

Electronic Farebox 'Quick Reference'

Standard Procedures

- Buses -

Logging on

1	◆ If you see a flashing red dot, press [key] green DUMP key ◆ Display should read NO DRIVE *Note:* If you swing on mid-trip and NO DRIVE isn't displayed, you must first log-off the previous operator	→ press [key] to continue
2	◆ Display should flash FARE SET ◆ Enter the proper fareset for the route	→ press [key] to continue
3	◆ Display should flash NO DRIVE ◆ Enter your badge number	→ press [key] to continue
4	◆ Display should flash SIGNCODE ◆ Enter the same signcode used for electronic destination sign	→ press [key] to continue
5	◆ Display should flash RUN ? ◆ Enter your run number	→ press [key] to continue

Your log-on is now complete, but double-check your input and make sure the farebox is operating properly!

6	◆ Press [key] and repeat to review each of your entries. ◆ Once you are sure there are no mistakes, press the [key] green DUMP key.
7	◆ The numeric display will read 0.00. ◆ Press [0] to test the coin drop mechanism -- you will hear a 'churning' noise.

Now you're ready to accept passengers!

Fareset Listing

1	→	**60¢ (Local)**
2	→	$1.00 (Zoned Local)
3	→	$1.50 (Express)
4	→	$2.00 (Express)
5	→	$2.25 (Express)
6	→	60¢ (Local - Subway passes accepted)

Use fareset 6 on Routes 39, 49, CT1(701), CT2(747), CT3(708) & 01 (between Mass Ave & Dudley only)

Run Numbers

Always enter your assigned run number. If there is no run number for the trip (as in extra trips, special work, etc.), enter 7 followed by your area number:

Albany	→	7223	Arborway	→	7122
Cabot	→	7123	Quincy	→	7128
Bennett	→	7125	Charlestwn	→	7126
Fellsway	→	7126	Lynn	→	7129

At the end of each half-trip, you must change the signcode on both:

✔ the electronic destination sign
✔ the farebox

Before accepting passengers for your next half-trip, do the following:

- ◆ Press [key] and repeat until the display flashes SIGNCODE
- ◆ Enter the signcode for the next half-trip
- ◆ Press the [key] green DUMP key

Remember: use signcodes from the *Electronic Destination Sign and Farebox Codes* pamphlets -- -- do not use route numbers!

Logging off

✔You must log off each time you finish operating a bus -- -- this includes swinging off a bus mid-trip

- ◆ Press the [key] green DUMP key to clear all remaining coins
- ◆ Press [key] -- then press [key] -- the farebox will make a 'warbling' noise
- ◆ Press [key] and repeat until the display flashes DRIVER ?
- ◆ Press [key] to clear the current badge number
- ◆ Press the [key] green DUMP key -- after a couple seconds, the display will read NO DRIVE

***That's it!* You're logged off.**

Electronic Farebox 'Quick Reference'

Passenger Classification

- Buses -

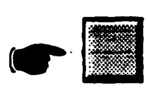
ELECTRONIC FAREBOX KEYPAD

The basics:

- If passenger pays correct fare **and** the farebox 'beeps,' -- DO NOTHING!
- If passenger swipes correct pass **and** the farebox 'beeps,' -- DO NOTHING!

but . . .

If neither of these things happen,
you must __classify the passenger__ by pressing the proper key from above!



- At the end of each half-trip, you must __change the signcode__!

__It's easy to remember__. . .

"Whenever you change the signcode on the electronic destination sign, enter the same signcode in the electronic farebox."

(*__Note:__* **If your bus doesn't have an electronic destination sign, you must still enter the proper signcode in the farebox after each half-trip.**)

Electronic Farebox 'Quick Reference'

Selected Examples of Fare Exceptions for Buses

(Partial List -- Other examples will be included on this sheet periodically)

What should I do if . . .

Zoned Local Routes (34E, 52, 62, 76, 134, 136, 137, 230, 238, 240, 350, 435)
Zoned Express Routes (305, 353, 426)
Zoned Limited-Stop Routes (436, 439, 441, 442, 450, 455, 458, 468)

I chose the fareset according to the highest fare on the route . . . but a passenger put in a lower fare because he/she is only going a short distance	I chose the fareset for the highest fare on the route . . . but a passenger used a lower-value pass because he/she is only going a short distance
☛	☛

Local Routes with Subway Transfers (01 between Mass Ave & Dudley, 39, 49)

A passenger gave me a transfer from the Orange or Green Line	A passenger put 85¢ in the farebox and wants a transfer to the Orange or Green Line
☛ 7	☛ 8

A student, senior citizen, or person with a disability paid for a transfer to the Orange or Green Line

☛

Routes 458/468

Fareset 5 ($2.25) is the highest fareset available, so when a passenger pays $2.50 for express service between Danvers and Haymarket, $0.25 remains on the numeric display

☛ the green DUMP key

All Routes

A passenger does not pay his/her full fare or does not pay any fare at all

☛

Remember:

☛*Classify your passengers on the farebox!*

Appendix C

Calendars for the Four Seasons

Table C.1 shows calendars for the four seasons (winters and springs of 1994 and 1996) with the reasons for eliminating specific days based on the symbols defined in Table C.2. Once again, since fareboxes were not probed daily in the winter and the early spring of 1996, the daily level data (including the control totals) for a probe day were divided equally between the two days covered by the probe. Some days with zero recorded revenue are used because the data for the following day are allocated back to the current day.

Table C.1: Calendars with reasons for eliminating specific days

Bartlett Garage Revenue (Daily Revenue + Unclassified Revenue) (Jan 94 - Mar 94)							
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Jan 94							
1 - 2	X	X	X	X	X	-	-
						HS	SE
3 - 9	23678.71	11185.32	14091.26	12478.87	12722.46	4451.14	-
	IE	W				S	S
10 - 16	20095.71	10252.2	13693.49	14089.44	14244.11	5548.55	-
	I	W			E	SE	SE
17 - 23	9073.26	15696.29	11615.27	12951.49	13178.08	5432.92	-
	IH	E				S	S
24 - 30	19751.08	14255.91	12648.47	12858.34	12889.95	6590.89	-
	I					S	S
31	19605.49	X	X	X	X	X	X
	I						
Feb 94							
1 - 6	X	14935.28	14096.44	13996.81	12588.39	7057.14	-
						S	S
7 - 13	18925.8	12234.16	10910.01	12878.66	12273.84	7079.04	-
	I		W			S	S
14 - 20	15575.26	15009.01	13482.54	15809.78	14234.67	6904.04	-
	I				E	SE	SE
21 - 27	11420.98	16587.08	12536.59	12861.48	13994.56	6590.47	-
	IH	E				S	S
28	16644.38	X	X	X	X	X	X
	I						
Mar 94							
1 - 6	X	16506.21	13674.04	10064.07	15280.6	7899.14	9.72
				W		S	S
7 - 13	19053.99	13085.1	13435.38	11800.41	13937.71	7181.76	-
	I					S	S
14 - 20	19615.32	12596.08	13191.06	11939.72	13880.8	6834.79	0.8
	I					S	S
21 - 27	18344.4	11548.55	13142.17	13354.36	14230.78	6746.71	-
	I					S	S
28 - 31	18706.42	11904.95	12790.24	12736.31	X	X	X
	I	U	U	U			

Bartlett Garage Revenue (Daily Revenue + Unclassified Revenue) (Apr 94 - Jun 94)							
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Apr 94							
1 - 3	X	X	X	X	17831.88	7378.93	-
					U	S	S
4 - 10	19701.63	13807.07	12049.37	12546.92	14422.14	6819.68	-
	I	U	U	U	U	S	S
11 - 17	18230.76	12756.8	12286.11	13563.52	14034.47	6549.06	-
	I				E	SE	SE
18 - 24	12745.19	13770.61	14078.85	15617.92	14461.3	7205.4	-
	IH	E				S	S
25 - 30	15595.75	14688.41	12686.52	12966.72	14192.53	7224.28	X
	I					S	
May 94							
1	X	X	X	X	X	X	-
							S
2 - 8	22613.08	13928.25	13149.61	11969.74	12725.61	-	-
	I					S	S
9 - 15	18474.41	13129.8	12155.21	12741.63	13430.16	-	-
	I					S	S
16 - 22	24333.31	11972.69	12795.42	12063.33	13944.41	7351.19	-
	I					S	S
23 - 29	14667.36	15750.18	12359.95	11097.35	14885.78	7131.69	-
	I				E	SE	SE
30 - 31	-	22767.03	X	X	X	X	X
	H	IE					
Jun 94							
1 - 5	X	X	14822.95	14426.15	14362.55	-	-
						S	S
6 - 12	24491.41	12822.98	13883.45	12846.35	13788.69	7710.56	-
	I					S	S
13 - 19	17067.18	12520.5	12764.5	13588.26	13094.84	6124.07	-
	I					S	S
20 - 26	16386.11	12899.49	13144.09	12909.08	12808.6	-	-
	I					S	S
27 - 30	25643.14	13308.44	12609.71	12513.79	X	X	X
	I	T	T	T			

Bartlett Garage Revenue (Daily Revenue + Unclassified Revenue) (Jan 96 - Mar 96)

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Jan 96							
1 - 7	-	34195.18	15393.46	63.95	26425.22	-	-
	T	I				S	S
8 - 14	20145.89	108.86	28486.39	208.4	24895.73	-	-
	I					S	S
15 - 21	19210.23	43.69	32017.37	0.02	22862.49	-	-
	IH	U				S	S
22 - 28	26951.8	39.89	23641.35	39.37	29287.73	-	1744.41
	I					S	S
29 - 31	25047.53	0.02	24979.87	X	X	X	X
	I						
Feb 96							
1 - 4	X	X	X	230.96	25081.56	-	-
						S	S
5 - 11	29233.94	874.35	22859.84	13.57	22660.45	-	-
	I					S	S
12 - 18	21679.27	1676.24	11347.73	-	31723.42	-	-
	I	A	A	A	A	S	S
19 - 25	17528.53	0.14	32411.6	129.36	22805.47	12715.33	-
	IH	E	E			S	S
26 - 29	17656.94	0.02	25414.14	-			
	I						
Mar 96							
1 - 3	X	X	X	X	24643.75	-	-
						S	S
4 - 10	28362.94	39.84	25385.29	24.34	22986.81	-	-
	I					S	S
11 - 17	27028.27	0.2	23261.16	-	27926.44	4329.73	-
	I					S	S
18 - 24	20365.15	89.16	25971.81	48.02	22032.39	11923.67	-
	I					S	S
25 - 31	18942.31	-	25743.03	12198	12668.21	8362.14	-
	I					S	S

Bartlett Garage Revenue (Daily Revenue + Unclassified Revenue) (Apr 96 - Jun 96)							
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Apr 96							
1 - 7	21371.07	63.66	27880.98	13292.3	11595.45	8397.91	-
	I					S	S
8 - 14	16650.65	20.1	20756.22	-	25462.38	9328.99	-
	I					S	S
15 - 21	13779.31	14504.53	11912.92	93.84	23595.9	9866.48	0.7
	IE	E				S	S
22 - 28	20368.37	12841.71	14030.35	12236.15	13366.69	1603.96	-
	I					S	S
29 - 30	-	26648.92					
	A	I					
May 96							
1 - 5	X	X	15156.83	14958.08	13492.97	8876.35	-
						S	S
6 - 12	19653.71	12846.12	13278.43	11861.32	11133.76	9547.08	-
	I					S	S
13 - 19	17250.09	13391.31	12115.53	12394.15	12321.17	7120.01	-
	I					S	S
20 - 26	16153.88	13861.06	15049.22	7761.7	21.85	15194.1	-
	I			E	E	S	S
27 - 31	-	31341.95	10397.24	17288.84	13384.26	X	X
	H	IE					
Jun 96							
1 - 2	X	X	X	X	X	8624	-
						S	S
3 - 9	19496.14	15373.64	12511.76	13490.24	14214.91	7678.21	-
	I					S	S
10 - 16	16286.86	12442.61	14169.03	12012.58	12432.63	6966.89	1255.44
	I		U	U	U	S	S
17 - 23	15578.19	13938.34	8792.61	15574.03	12128.83	-	-
	U	U	U	U	U	- S	S
24 - 30	22791.64	1171.16	24137.6	-	25380.47	-	-
	T	T	T	T	T	T	T

Table C.2: Explanation of Symbols Used in Table C.1

N = No information on whether probe was performed on the previous day

I = Inflated revenue (due to the absence or insufficiency of probe on the previous day)

S = Saturday or Sunday

H = Holiday

E = Holiday Effect (a weekday near a holiday which has revenue significantly different from a typical weekday)

W = Inclement Weather (revenue significantly different from a typical weekday)

U = Unused Day (in order to balance the number of days in each timetable)

T = New Timetable (not in the timetables of interest)

- = No available data (no probe performed on that day)

X = Non-existing day on the calendar

Any day with one or more of the above symbols is not used for the analysis

References

[1] United States Department of Transportation, "*Transit Data Collection Design Manual*", June 1985.

[2] United States Department of Transportation, Federal Transit Administration, "*Advanced Public Transportation Systems: The State of the Art Update '96*", January 1996.

[3] Robert Fellows, "*Using and Enhancing Automatic Vehicle Identification to Improve Service Control on the MBTA Green Line*", Master of Science in Transportation Thesis, Massachusetts Institute of Technology, September 1990.

[4] Canadian Urban Transit Association, "*Automatic Vehicle Location and Control Workshop Proceedings*". November 15, 1987.

[5] Massachusetts Bay Transportation Authority, "*Farebox Repair, Maintenance, & Service Manual M00511-3SM*".

[6] James Mulqueeny Jr., Sarah Labelle, Ross T. Patronsky, Joseph Simonetti, "*What to Do With Your New Electronic Farebox Data*", Chicago Transit Authority, Presented at TRB Conference on New Public Transportation Technology, San Francisco, August 1992.

[7] Nigel Wilson, Peter Furth, MacDorman & Associates, "*Review of MBTA Data Collection Plan*", Design of Service Quality Measures and Planning Standards, prepared for the Massachusetts Bay Transportation Authority by the Conservation Law Foundation, Inc., July 11, 1995.

[8] Transportation Research Board National Research Council, "*Transit Cooperative Research Program Synthesis 10*", page 44, sponsored by the Federal Transit Administration, 1995.

[9] Massachusetts Bay Transportation Authority (MBTA) Homepage: http://www.mbta.com

[10] Alan Castaline and Clinton Bench, "*Memorandum on Operator Compliance with Electronic Farebox Data Entry Guidelines*", April 18, 1996.

[11] Thomas C. Dawson, Dennis Ryan, "*Integrating Electronic Fareboxes at the CTA*", Chicago Transit Authority, December 1993.

[12] Federal Transit Administration (FTA) Homepage: http://www.fta.dot.gov