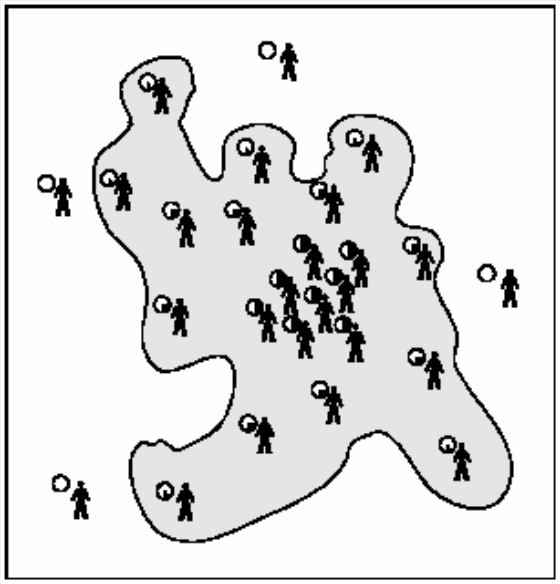
Ridership Models at the Stop Level



Final Report

BC137-31

National Center for Transit Research
Center for Urban Transportation Research
University of South Florida

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NCTR-473-04, BC137-31		

4. Title and Subtitle	5. Report Date
Ridership Models at the Stop Level	December 2004
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Xuehao Chu	
9. Performing Organization Name and Address	10. Work Unit No.
National Center for Transit Research (NCTR) University of South Florida 4202 E Fowler Av, CUT 100, Tampa, FL 33620-5375	11. Contract or Grant No. DTRS98-G-0032
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Office of Research and Special Programs U.S. Department of Transportation, Washington, D.C. 20590 Florida Department of Transportation 605 Suwannee Street, MS 26, Tallahassee, FL 32399	14. Sponsoring Agency Code

15. Supplementary Notes

Supported by a grant from the Florida Department of Transportation and the U.S. Department of Transportation

16. Abstract

This report documents the development of a ridership model at the stop level for Transit Level of Service (TLOS), a software package of the Florida Department of Transportation (FDOT) for measuring transit availability and mobility. This report is in two separate parts. This first part discusses issues in formulating a ridership model on several aspects, including literature review, the study area, analytical framework, data needs, and model-formulation issues. For each issue, the discussion focuses on the alternative approaches to dealing with the issue, any resolution, and CUTR's recommendation, if different from the resolution. The second part discusses the data made available by Kittelson and Associates, Inc. under a separate contract with the FDOT, the data used for model estimation, and the selection of a final model. The final model relates average weekday boarding at a TLOS stop with six categories of factors: 1) socio-demographics in a catchment area; 2) TLOS value; 3) the street environment for pedestrians; 4) accessibility to population and employment; 5) interaction with other modes; and 6) competition with other TLOS stops. Despite some data limitations, the model fits the data well and behaves as expected.

17. Key Words	18. Distribution Statement
TLOS, FDOT, boarding, weekday, accessibility, stops	Available to the public through the National Technical Information Service (NTIS), 5285 Port Royal Road, Springfield, VA 22161, 703-487-465, and through the NCTR website at http://www.nctr.usf.edu

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	64	

TABLE OF CONTENTS

INTRODUCTION 1

PART I: ISSUES 2

LITERATURE REVIEW 2

FRAMEWORK 7

DATA NEEDS 13

MODEL FORMULATION ISSUES 31

PART II: EMPIRICAL WORK 37

ESTIMATION DATA 37

ESTIMATION 39

SUMMARY 44

ACKNOWLEDGMENT 44

REFERENCES 44

APPENDIX: SELECTED TEST RUNS 47

LIMDEP FILES 47

OTHER FILES 48

SATURDAY RUNS 48

WEEKDAY RUNS 54

LIST OF TABLES

Table 1. Directional Daily Activities at Selected Stops, Route 140, LeeTran 10
Table 2. Quality of Service Factors and their Treatment 16
Table 3. Distribution of Stop Activities, LeeTran..... 33
Table 4. Data Available 38
Table 4. Poisson Regression of Weekday Total Boarding at TLOS Stops 43

LIST OF FIGURES

Figure 1. Framework for Identifying Potential Predictors..... 15
Figure 2. Zones, Sectors, and Sub-sectors for Crime Data..... 19

INTRODUCTION

The Florida Department of Transportation (FDOT) has developed Transit Level of Service (TLOS), a software package that measures transit availability and mobility (Kittelson and Associates and URS, 2000). Versions 1 and 2 of the software focused on transit availability in terms of percept person-minutes served for individual stops. This measure of transit availability is also called the TLOS indicator or simply TLOS as well. Version 3 expanded the previous versions by including transit mobility in terms of how much service is provided between any two stops and how long it would take to make the trip.

The TLOS indicator makes two advances over the conventional measures of transit availability (Ryus et al., 2000). The Transit Capacity and Quality of Service Manual (TCQSM) addresses transit availability by three separate measures--service frequency, hours of service, and spatial service courage (TRB, 2003). The first advance of the TLOS indicator is that it implements the conventional measures at the operations level. It allows more detailed analysis of transit availability. It can be calculated at a wide range of geographical levels--bus stop, route, corridor, neighborhood, or system—and of temporal levels—15 minutes, 1 hour, 1 day, or 1 week. The second advance of the TLOS indicator is that it combines the three aspects of transit availability—service frequency, hours of service, and spatial coverage—into a single measure.

The Department has contracted with both the Center for Urban Transportation Research (CUTR) and Kittelson & Associates, Inc. (KAI) to incorporate in Version 4 a predictive tool that would determine how many people may board transit vehicles at individual stops as a function of the TLOS indicator. KAI searched and reviewed the literature and provided both the review and the raw documents to CUTR. KAI also collected the required data and provided CUTR with the data electronically. CUTR's responsibility through this project was model estimation. Finally, KAI was going to incorporate the final stop-level ridership model into TLOS Version 4.

Specifically, the objective of the CUTR effort was to develop the model for ridership prediction at individual stops within the constraints of four criteria. It should address fixed-route and guideway transit services provided by the Jacksonville Transit Authority within a single model of stop-level ridership. It should consider boardings by direction. It should consider boardings on a daily basis. It should consider park-and-ride lots in ridership estimation but should not develop a separate demand model for using park-n-ride lots. Finally, the development of this model shall be limited to input variables for which data are provided by KAI.

This report documents the CUTR effort. Specifically, it proposes and carries out an indirect application of the TLOS indicator where the TLOS results are used in combination with other data. It hypothesizes that the TLOS indicator is related to transit patronage in a particular way and tests this relationship using data from Jacksonville, Florida. This relationship can be used to estimate potential changes in patronage as a result of changes in transit availability. The report is in two parts. The first part discusses issues considered in formulating the empirical work, and the second part presents the empirical work. An appendix contains the results of selected test runs using LIMDEP as well as simple instructions on using the data and the script file contained in the CD attached to this report.

PART I: ISSUES

This part discusses issues considered in formulating the empirical work reported in the second part. Several areas are considered, including a literature review, the study area, an analytical framework, data needs, and model-formulation issues. For each issue, the discussion focuses on the alternative approaches to dealing with the issue, any resolution, and CUTR's recommendation, if different from the resolution.

LITERATURE REVIEW

Transit agencies need to know the impacts of changes to their network or service levels on transit patronage. Some of these changes are modifications to existing services while others represent proposed new services. Transit agencies need to know these impacts for a variety of reasons (Multisystems 1982). There are always competing demands for adding new services or maintaining existing services. Information on these impacts provides a basis on which to allocate resources. Transit agencies need to know these impacts in order to prepare budget requests for proposed new service plans to their governing board. In this case, cost and revenue forecasting must be reasonably accurate. Finally, these impacts are important inputs into detailed network and scheduling planning, particularly for new service plans.

The traditional four-step process is ineffective for assessing such impacts. First there are accuracy and relevancy reasons (Multisystems 1982). The four-step process was designed as a planning tool for large-scale, capital intensive changes. The unit of analysis is typically at the zonal level. Errors from this process may be larger than the impacts of low-cost, operational changes. In addition, low-cost transit changes typically have little impact on the diversion of automobile trips to transit. While new ridership may be generated, it will come mainly from people making new trips and diverting from walking or biking. The increased ridership will have virtually no impact on the highway level of service in most cases. Besides accuracy and relevancy, the traditional four-step process lacks flexibility in the possible changes to a transit network or services provided that a transit agency can explore and implement.

While simplified versions of the traditional four-step process (Horowitz 1984; Horowitz and Metzger 1985) have been developed for specifically accessing these impacts, most previous work has focused on direct demand models at the route-level (Alperovich et al. 1977; Cherwony and Polin 1977; Kemp 1981; Menhard and Ruprecht 1983; Stopher and Mulhall 1992; Hartgen and Horner 1997). While these route-level models avoid some of the problems with the traditional four-step process, they have their own difficulties (Alperovich et al. 1977). They assume homogeneous service levels and land use along each route. This assumption is particularly problematic along long routes that start at the central business district and go into deep suburbs. They are ill equipped to deal with inter-relationships in a transit network that occur at the sub-route level.

Modeling and forecasting transit patronage has recently been advanced to the segment-level (Peng et al. 1997; Kimpel et al. 2000). Segments may be defined by time-point stops as by Kimpel et al. (2000) or by fare zones as by Peng et al. (1997). This advance recognizes the spatial variation of patronage and service supply across the segments of a route. It has been

aided by geographic information systems (Peng and Dueker 1995) as well as by the availability of patronage data at the segment level from automated passenger counters. It has allowed the assessment of new policy instruments such as service reliability (Kimpel et al. 2000). It has also gained new insights into inter-relationships in a transit network and their effects on patronage that traditional route-level analyses were unable to provide (Peng et al. 1997).

The segment-level models still share some of the difficulties of route-level models (Strathman et al. 1997; Peng 1997; Kimpel et al. 2000). The social-demographics along a route segment are assumed to be homogeneous. It is difficult for segment-level models to take into account important stop-level characteristics such as pedestrian access, stop amenities, and special generators. It is difficult to represent the competing effects in a transit network at the segment level. It is difficult to represent some service variables at the segment level. It is also necessary to address the simultaneous determination of both patronage and service levels. Many of these limitations result because transit patronage is analyzed at the route-segment level but is realized at the stop level.

Strathman et al. (1997) argue that patronage models for service planning should be estimated at the route or segment level but not at the stop level. They argue that service planning is most concerned about accessing the impacts of change in service levels on patronage, and that service levels are usually planned at the route or route-segment level. This argument takes too narrow a view of service planning, which is far more than planning for frequency changes. This argument also takes too narrow a view of the potential patronage impacts of changes in service levels. Suppose that the frequency is increased along route A that intersects route B. Only at the stop level would one be able to capture the change in transfer wait time for a passenger transferring from route B to route A at their intersecting point. It may be true that the most relevant level of patronage impacts from service changes is at the route or segment level. That does not necessarily mean, however, that patronage models should also be estimated at the same level.

The only previous attempt to modeling and forecasting patronage at the stop level is by Kikuchi and Millkovic (2001). They use fuzzy inference and directionless stops. This approach is similar to the traditional cross-classification approach to trip generation modeling (Meyer and Miller 2001), with the boundaries of the discrete classes being fuzzy. While it does serve as an alternative to other methods, there is little advantage over traditional regression-based methods. In fact, the sensitivity of predictions to changes in continuous factors influencing patronage is limited as a result of using them in few discrete categories.

STUDY AREA

This section discusses the unique characteristics of the study area that are likely to affect several aspects of the research design, including sampling, model structure, and model transferability. The FDOT has already selected Jacksonville as the study area (Ausman, 2001). That means that the ridership model will be developed with data collected from Jacksonville. The selection was based on the fact that the transit system in Jacksonville (JTA) has automatic passenger counters (APCs) on some of its transit buses.

Passenger Data

JTA had a total of 30 APCs. Five of these APCs were permanently installed on the trolley. These APCs allowed a complete count of stop activities on the trolley. Four of these were used to collect data for the National Transit Database. The other 21 APCs were installed on fixed-route buses that rotated across routes and days such that at least one one-way bus trip is counted once every month for weekdays, Saturdays, and Sundays, respectively.

According to JTA, its APCs achieve a level of accuracy of over 95 percent for counting stop activities. In addition, alighting is likely to be more accurately counted than boarding because passengers are more likely to bunch when boarding than when alighting. The APCs provide separate data on boarding and alighting. Additional data are available on running time, etc. The APCs do not separate direct boarding from transfer boarding.

The people mover has no APCs. Stop activities there would come from the station turnstiles. The incoming and outgoing volumes through the turnstiles would be used as the boarding and alighting volumes for each stop.

Stop Density

The density of stops affects both the delineation of quarter-mile stop buffers and the predictability of activities at close-by stops. No overlapping between stop buffers occurs on average if average density is three stops per mile. On the other hand, stop buffers completely overlap with each other on average if the average density is five stops per mile. The 1999 National Transit Database indicates that the JTA bus system had about 1,000 directional route miles. According to JTA, its bus system has about 5,000 stops that are spatially unique after taking into account overlapping route miles. As a result, JTA's bus system has an average of five stops per route mile.

There are at least three consequences of dealing with a high density of stops. The first is the cost of data collection. This consequence is no longer an issue because the KAI contract covers data collection for all stops. The second consequence is that there may be no measurable variable that can explain differences in stop activities between close-by stops. For highly overlapped stops, the only potential variables to explain differential stop activities are stop amenities. Frequently, however, those highly overlapped stops are identically simple: one metal pole but nothing else. Even when amenities such as a shelter are present, we may not be able to include them in the model because of the statistical problems that doing so would bring. The third consequence is the measurement of buffer characteristics when buffers overlap.

There are two potential approaches to dealing with this issue. First, a statistical approach to avoid these consequences is to use a random sample of the stops. We could use all stops for the trolley and people mover, but only a simple random sample of the bus stops or a stratified sample with the quadrants as the strata. As discussed above, quarter-mile buffers do not overlap on average when the density is no more than 3 stops per route mile. A reasonable sample size may be 2 stops per route mile. This would mean a total of about 2,000 stops for model development. Second, the alternative is to exclude overlapped buffer areas. The rationale is to

avoid the multiple counting of characteristics. Multiple counting would not be a problem if all measured characteristics are not cumulative. For this project, however, some of the buffer characteristics are likely to be cumulative, including the TLOS measure, the number of people, and number of jobs within individual buffers. For the purpose of measuring accessibility to population and employment, however, all stops still need to be considered with overlapped buffer areas excluded.

The Department seems to prefer the partial buffer approach. CUTR suggests that the statistical approach be taken. Even if the statistical approach is taken for model development, however, the partial buffer approach would still need to be used for model application when stops are dense and all need to be considered.

Unique Characteristics

Certain characteristics of the study area may affect whether, and the degree to which, the model is transferable to other systems. Some of these are differences among transit systems that can potentially be accounted for through the models. Differences in modes available and fare policies are highlighted. Others are regional differences that cannot be accounted for through the ridership models at the stop level either because data for these stops are from a single region.

System Differences

One system difference is in the modes available. Differences in the modal mix go both ways. First, the JTA system includes three modes of fixed-route transit service: bus, the people mover, and the trolley. Most other systems in Florida do not have a people mover or trolley. Second, the MTDA system in Miami-Dade County has both heavy rail and bus rapid transit that the JTA system does not.

The transferability of models developed with JTA data depends on how the models are eventually specified and the success of these specifications. For the people mover or trolley, either mode-specific models or a single model with mode-dummies would allow transferability. For heavy rail or bus rapid transit, however, neither would work because they are unavailable in JTA.

One potential solution of the transferability issue would be our ability to measure and include modal characteristics that are universal to all of these modes. The various transit technologies differ not in whether one technology has certain characteristic but others do not. Rather they differ in the values of a set of characteristics that are common to all of these transit technologies. This idea is based on the so-called abstract-mode theory (Quandt, 1970).

The other system difference considered is in fare policy. The trolley is free for everyone. The buses and people movers are free for seniors, and a flat fare is assessed for any other group of riders. JTA does not offer fare transfers either for intra-modal or inter-modal transfers.

Since the ridership models will be area-based rather than population-based, the mix of fare policies across user groups prevents the models from being fully transferable to other areas that have different mixes of fare policies. The degree of transferability depends on whether some of the fare policies are part of the model. A total exclusion of fare policies in the models would significantly reduce their transferability. This is particularly true for bus services. Unlike rail services, bus riders in general are more sensitive to fare policies. Inclusion of fare policies in any form in the models would greatly increase their transferability to other systems, at least in terms of fare policies.

Fare policies may enter the models in different forms. First, the base fare may be directly used in the model. While it is constant across the stops of the same mode, it does vary across modes. The variation, however, will be limited because there are few stops on the trolley or the people mover. Furthermore, whether the base flat fare can be directly used in the model depends on whether the different modes are modeled separately. Second, an average fare in the buffer area around any given stop can potentially be used directly in the model. In order to do this, however, we would need to know the mix of the population in the buffer area who pay different fares. Third, the base fare may be indirectly used in the model through a fare-based measure of accessibility from any given stop to downstream population and jobs and a similar measure from upstream population and jobs to any given stop. The variation in the resulting accessibility levels across stops results from the fact that transfers are not free. The good thing about fares is that they are not likely to be endogenous to stop activities. In contrast, some of the supply characteristics are.

Regional Differences

There are potentially many regional differences that can influence the overall ridership in a region, but cannot be adequately controlled in ridership models at the stop-level from a single region. Examples include the dominance of a university such as in Gainesville, and the dominance of tourism such as in Orlando. Some of these regional differences may be controlled to some degree in stop-level models. One approach would be to measure land use more specifically than by the number of jobs or job density. At a minimum, industries could be specified. For example, workers in service industries may have a different propensity to use bus transit than workers in information industries. The Jacksonville area appears to have a younger population than many other areas in Florida. In 2000, seniors, persons who age 65 or older, accounted for about 18 percent of the state population. In Jacksonville, they accounted for only 10 percent. As a result, the model should adequately control for the age distribution in the buffer areas.

Resolution

The Department wants to consider all three modes. CUTR recommends the following: 1) Fares be considered along with travel time and the number of transfers in computing the downstream number of jobs and people reachable from a stop and the upstream number of jobs and people reachable to a stop; 2) If the fare- and transfer-based measures are highly correlated, use the fare-based measure; and 3) Fares be considered as a direct predictor in models that combine all three modes.

Temporal Gaps in Data

Ideally, all variables used in developing a demand model need to be collected for the same temporal period. This may not be feasible with the resources available in this case. The gap results from the fact that data on land use and population characteristics will come from existing sources that reflect pre-2001 conditions. Data on stop activities will be collected later this year that reflect averages of 2001 conditions. Data on transit schedules will reflect current conditions, which may differ from those earlier this year.

While we may not be able to match land use data with others, we can match the timing of stop activities with transit supply data. This may be done retrospectively by selecting a period in which supply characteristics were not changed and for which data for both stop activities and supply characteristics would be measured. In order to precisely define the period for data collection, information on changes in supply characteristics at JTA for 2001 would be needed. Alternatively, a future period may be selected and related data could be collected for that period.

FRAMEWORK

We consider several elements of a potential analytical framework. These include the basic concept, measurement of stop activities, stop types, and model types.

Concept

According to the CUTR scope of services for this project, we are interested in stop activities: boarding and alighting. That is, we are not interested in where these people will go once onboard or where they came from. That is, we do not consider the destinations of boarding or the origins of alighting. As a result, we will not have information on the characteristics of the stop-to-stop trips.

What we will have are the characteristics of the stop area, the characteristics of transit services available at the stop, and how this stop relates to the entire transit network. The characteristics of the stop area include at least three components: characteristics of the stop itself, the relationship between the stop and the roadways, and the characteristics of the surrounding areas of the stop. The characteristics of transit services include reliability at the stop as well as frequency and service span at the route level. The relationship of the stop with the entire transit network includes at least two components: the accessibility of the stop to downstream land use, the accessibility of upstream land use to the stop, and the presence and absence of both competing and complementary routes at and near the stop.

The closest concept used in regional transportation modeling is trip generation and attraction rates at the zonal level. In this case, the trip generation and attraction rates are at the stop level and for transit only.

Modeling origin-destination pairs is in the Department's longer-term vision for TLOS. In order for the tool to predict origin-destination pairs would require data on origin-destination travel

patterns. One may model origin-destination pairs either indirectly through considering both trip generation and distribution, or directly through considering transit trip flows between stop pairs. FSUTMS uses the indirect approach. The direct approach might be more promising for TLOS. In fact, Thompson (1997) uses the direct approach to model zone-to-zone transit flow with on-board survey data from Sacramento. Data on origin-destination pairs would be used for validation in the indirect approach and for calibration in the direct approach.

While the current version will not specifically model origin-destination pairs, it will have a number of elements related to the longer-term vision. Destinations will be taken into account indirectly by the number of people and jobs accessible from a transit stop within a given travel time, and within a given number of transfers. TLOS 4 can also provide zone-to-zone travel times, which could be used to develop a weighted travel time between a stop and all potential destinations within a given travel time.

It was determined to focus on stop activities. When possible, make the model developed consistent with future versions of TLOS that would have the capability of predicting transit trips between stop pairs.

Measurement of Stop Activities

The measure of stop activities involves a number of issues, including measurement by direction, temporal aggregation, measurement by day of week, boarding versus alighting, direct versus transfer activities, and total versus per capita activities.

Directional Measurement

The first issue is how we should treat transit activities at the two stops in opposite directions of the street. One approach would predict stop activities for all spatially unique stops. It is useful to know stop activities by direction for certain operations planning. Suppose, for example, that boarding is extremely directional at a particular street location. The transit agency may want to place a shelter at the stop with large boarding volumes but not at the other stop. The challenge is to find measurable attributes that can account for typically significant differences in boarding activities between the two stops in opposite directions in the same general location. The characteristics of the buffer areas would not help because they are likely to be similar for these two stops.

The alternative is to predict transit activities by street location. That is, each pair of stops on the opposite sides of a road at the same general location would be combined for prediction purposes. The combined approach would make model development and prediction of transit activities relatively easier. Once combined, one no longer needs to account for the typically significant differences in activities between the two directions. One challenge is that this approach would not work with routes on a loop. One way to deal with loops would be to model them separately. Another challenge is that the two opposite stops often are not right across from each other but offset by some distance.

PBQD et al. (1996a) used daily boarding with the two directions combined in estimating stop-level ridership models for light rail and commuter rail. The combined approach was taken because some transit agencies could not provide daily boarding by direction for individual stations.

FDOT has decided that stop activities be measured for individual stops that are physically separated from each other.

Temporal Aggregation

The second issue is the level of temporal aggregation at which we should measure stop activities. Arguments can be made for measuring stop activities at less aggregated than daily levels. First, TLOS measures percent person-minute served for 15-minute intervals and travel time between pairs of stops for 2-hour intervals. Second, stop activities are much higher in peak periods than other times. The challenge to model stop activities at such fine levels is that standard variables would not be able to capture the underlying factors for such temporal differences in stop activities. One approach to this challenge would be to model stop activities separately for different periods of the day as done by Kimpel et al. (2000). FDOT has decided that stop activities be measured on a daily basis for TLOS 4.

Day of Week

The third issue is whether daily stop activities should be measured separately between weekdays and weekend days. Differences in the amount of services provided between weekdays and weekend days are not a reason for resolving this issue. Such differences can be reasonably accounted for by including certain supply characteristics at the route level such as frequency and service span. A reasonable argument for separating weekdays from weekend days is that the activity space differs dramatically between weekdays and weekend days. Most people do not work on weekend days. Many businesses and public agencies are either closed or have reduced operating hours.

To account for differences in the activity space, one of three approaches may be used. One is to use dummy variables for Saturday and Sunday. Another one is to model weekend days separately. A third approach would account for the temporal distribution of opportunities for activity over the different days of a week. This distribution would be approximated by the distribution of non-transit trips in Jacksonville or Florida based on the 1995 Nationwide Personal Transportation Survey. This distribution would be constant across the routes while in real life it is likely to vary. This information may enter the model as a separate variable or interact with other variables that measure opportunities for activity. The Department believes that weekdays and weekend days should be separately modeled.

Barding versus Alighting

The fourth issue is how we should treat boarding and alighting. Three options are available in dealing with boarding and alighting: boarding only, boarding and alighting separately, and combined total. It is arguable that either of these may be used.

In an ideal world, one would intuitively think that the number of boarding in one direction would be equal to the number of alighting in the other direction for a given destination. Under this ideal world, knowing boarding means knowing both alighting and the sum of both. PBQD et al. (1996a) modeled boarding counts only (daily with the two directions combined) at the stop level for light rail and commuter rail.

In the real world, however, daily boarding and alighting can differ for a given destination for a number of reasons such as one-way trips, competing routes, overnight stays, and circular routes. Table 1 below shows the on and off activities at five stops of Route 140 of LeeTran (CUTR, 2001). Boarding and alighting match extremely well at Edison and Crystal stops, while the others do not. It appears that boarding and alighting in the two opposite directions at the same location tend to be close for stops serving a single route. They are more likely to deviate from each other at more complex locations.

Table 1. Directional Daily Activities at Selected Stops, Route 140, LeeTran

Destination	Northbound		Southbound	
	On	Off	On	Off
Heitman	0	0	2	10
Monroe Station	73	177	226	65
Edison	5	25	26	6
Edison Mall	109	98	108	131
Crystal	18	4	5	17

In addition to these observed differences in boarding and alighting volumes for individual destinations, boarding and alighting may be determined through different sets of factors. For example, some of the factors that are important to boarding are not so to alighting, particularly those related to stop amenities. Also, boarding is what matters most for certain destinations (likely to be residential) while alighting is what matters most for other destinations (activity centers).

The arguments made under separate considerations above would go against combining boarding and alighting. However, the combined treatment in a study of BART stops seems to be quite successful in terms of the amount of variations in total stop activities explained by a set of standard explanatory variables (PBQD et al., 1996a).

If travel directions were combined, daily boarding and alighting would be the same in most cases at the individual stop level. If directions are separately considered, however, they differ in most cases.

Conceptually, modeling boarding differs from modeling alighting. Boarding is trip generation, while alighting is trip attraction.

The consensus of the participants at the Jacksonville kickoff meeting seems to be that it would be useful to know boarding and alighting volumes separately. CUTR agree with the consensus.

However, CUTR argues that it is not necessary to predict alighting separately. It is true that alighting may not be always similar to boarding in the other direction for a given location. However, these quantities are likely to be the same magnitude. In addition, the usefulness of alighting volumes is reduced once the measurement is on a daily basis and the two directions are combined.

Direct versus Transfer Activities

The fourth issue is whether we should model direct stop activities separately from transfer activities. The only argument against separating them is to reduce the number of separate models. The argument for separating them is much stronger. That is, the potential predictors of direct stop activities are expected to be somewhat different from those of transfer activities. For modeling purposes, therefore, they should be separated. This argument is even stronger in the Jacksonville case because as many as half of its transit users transfer at least once on a daily basis. In estimating a route-level ridership model, Kemp (1981) separates transfer and non-transfer boarding as two endogenous variables with a simultaneous-equations approach. It is determined that direct and transfer activities will be combined. The APC data on stop activities do not distinguish these two activities.

Total versus Per Capita Activities

The fifth issue is whether stop activities should be expressed in terms of total daily, per capita, or per acre of buffer area. In addition to modeling total stop activities, the BART study (PBQD et al., 1996a) also modeled total activities per capita and total activities per acre. According to the authors, expressing activities on a per capita or per unit of geographic area provides models that have the most transferability to other systems. Either alternative does not seem to be appropriate for this project. Since the sizes of the buffer areas vary little, and will be identical if full buffers are used, expressing activities as total daily and per capita makes no difference for modeling purposes. Since we are not going to limit ourselves to home-based boarding only, stop activities per capita would not make much sense for stops in areas with large commercial activities. It is recommended to use total daily activities.

Stop Types

The issue is how we should treat the various stop types in a transit system. The discussion below includes classification, modeling options, advantages, and feasibility.

Stops may be grouped by different criteria for modeling purposes. Below are several examples:

- By mode: fixed-route bus, trolley, and people mover.
- By route type: interliner, radial, cross-town, connector, express, looper, feeder, etc.
- By direction: single direction and both directions.
- By schedule: regular and time-point.
- By position: terminal and middle.
- By access: walk, bike, transit, auto (park-n-ride lots).

- By inter-route relationship: transfer and non-transfer; single route and multiple routes; timed-transfer and non-timed transfer.

There are three options for considering stop types. One option is to capture differences across stop types with a set of supply characteristics. These would include average route speed, frequency, service span, stop reliability, upstream accessibility, downstream accessibility, the number of transfers, etc. For example, interliners allow rides from one suburb to another to avoid a transfer and fare between the two original routes. Also time-point stops are likely to offer more reliable services than other stops. In addition, terminal stops do not provide downstream accessibility without a transfer. The second option is to use dummy variables to indicate whether a particular stop belongs to a certain type. A dummy variable takes 1 if a stop belongs to the particular group and 0 otherwise. The dummy variables are used to capture differences in stop activities cross stop types that are not captured by other predictors. The second option may be used with the first option so that the dummy variable could capture additional differences that the included supply characteristics could not. The third option is to model stop types separately. That is, each stop type under a given classification has its own model.

Each option has its own advantages. The first option allows a more compact specification of the models. If we were able to characterize them in terms of a limited set of supply characteristics, we would only need a single model. One advantage of other two options is the likely increase in model transferability. This would be realized only when different stop types couldn't be adequately characterized with the limited set of supply characteristics. One additional advantage of separately modeling certain stop types over using dummies is that different sets of predictors may be used for different stop types. It is possible that stop activities on radial routes may have a quite different set of predictors from those on cross-town routes. One obvious disadvantage of modeling stop types separately is the increased number of models to keep track. It is likely that we may need to use different options for different stop types. The feasibility of using each option varies. The use of dummies is feasible in most cases.

Whether we can use the first option depends on whether we can differentiate some of the stop types by the supply characteristics included in the models. Timed transfer points tend to reduce transfer waiting, which is unlikely to be a separate characteristic in addition to route frequency. Also, park-n-ride lots are intended to draw potential riders who access transit by the automobile from a larger area than regular stop buffers. None of the supply characteristics would be able to capture the unique features of park-n-ride lots. In addition, our ability to use the first option will be limited due to the difficulty of characterizing stops with route information because many stops are served by multiple routes.

When separate modeling is desired, its feasibility depends on whether enough stops exist within each type. In the case of modal classification, the JTA trolley and people mover have 12 and 8 stops, respectively, which are not enough to model the two modes separate from the bus mode.

Sometimes none of these options may be feasible. Consider the stop where two radial routes used to intersect and riders from one route could transfer to the other. Suppose that these two routes are now linked to become a single interliner. The boarding and alighting activities at this stop from transferring between these two routes are suddenly vanished. By including in the

model fare, transfers, and speed, we are likely to capture the increased attractiveness of both routes to riders who made transfers between these two routes. However, there seems to be nothing we can do to the model that would make the predicted boarding and alighting at this stop sensitive to such structural changes in the network. (These two routes were complementary to each other. Once combined into an interliner, they become a single route. There may be something to be measured there.)

It is determined to consider the various stop types. It is to be determined whether dummies should be used or they should be modeled separately.

Model Types

A predictive model can be either causal or non-causal. A causal model is developed to explain a phenomenon while a non-causal model is not. A demand model is a causal model. However, a model that includes any variable that shows high correlation with demand is non-causal. A causal model is the preferred choice in general. Which of these two model types is used in practice depends both on the availability of a theory behind the model and the availability of data. A predictive model without a theory behind it is frequently non-causal. Even with a sound theory, a predictive model without adequate data available is frequently non-causal too. In this particular case, a non-causal model is likely to be the end result because we do not appear to have a sound theory on generation or attraction.

Should we model transit activities directly or model total travel activities and modal split? As in the traditional four-step process for regional transportation demand forecasting, stop-level activities may be predicted in one of two alternative approaches. One approach involves two steps: 1) predict total travel activities by all modes, and 2) split this total into transit and other modes. Alternatively, one could predict transit activities directly. The selection involves at least two issues. One issue is which approach would result in smaller prediction errors. CUTR is unaware of any study that has looked into this issue at any geographical level. The other issue is whether data are available on total travel activities and modal split for individual stop buffers. The literature review by KAI suggests the four-step process as an alternative approach to modeling transit ridership (KAI, 2001). The Department has chosen the approach of direct transit generation and attraction. One reason for the choice is the availability of data on estimated transit boarding and alighting for individual stops.

DATA NEEDS

Before data collection and modeling, a set of potential predictors for stop activities were developed with the following assumptions:

1. Transfer and direct stop activities would be combined;
2. The models would be non-causal ones;
3. The direct-modeling approach would be used;
4. Model activities at individual stops;
5. Model boarding only; and
6. Park-n-ride lots would not be directly modeled.

The first three assumptions resulted from the constraints the research teams faced. For example, data were unavailable for separating transfer and direct stop activities. Also, resource constraints prevented data from being collected according to a causal model. The modeling of activities by stop pairs was part of the Department's vision for TLOS. However, there were no good data on transit trips by stop pairs. The fifth assumption resulted from the fact that boarding in one direction would be similar in magnitude to alighting in the other direction at the same location. The last assumption was made because the Department has a separate procedure for estimating demand for park-n-ride lots.

Framework

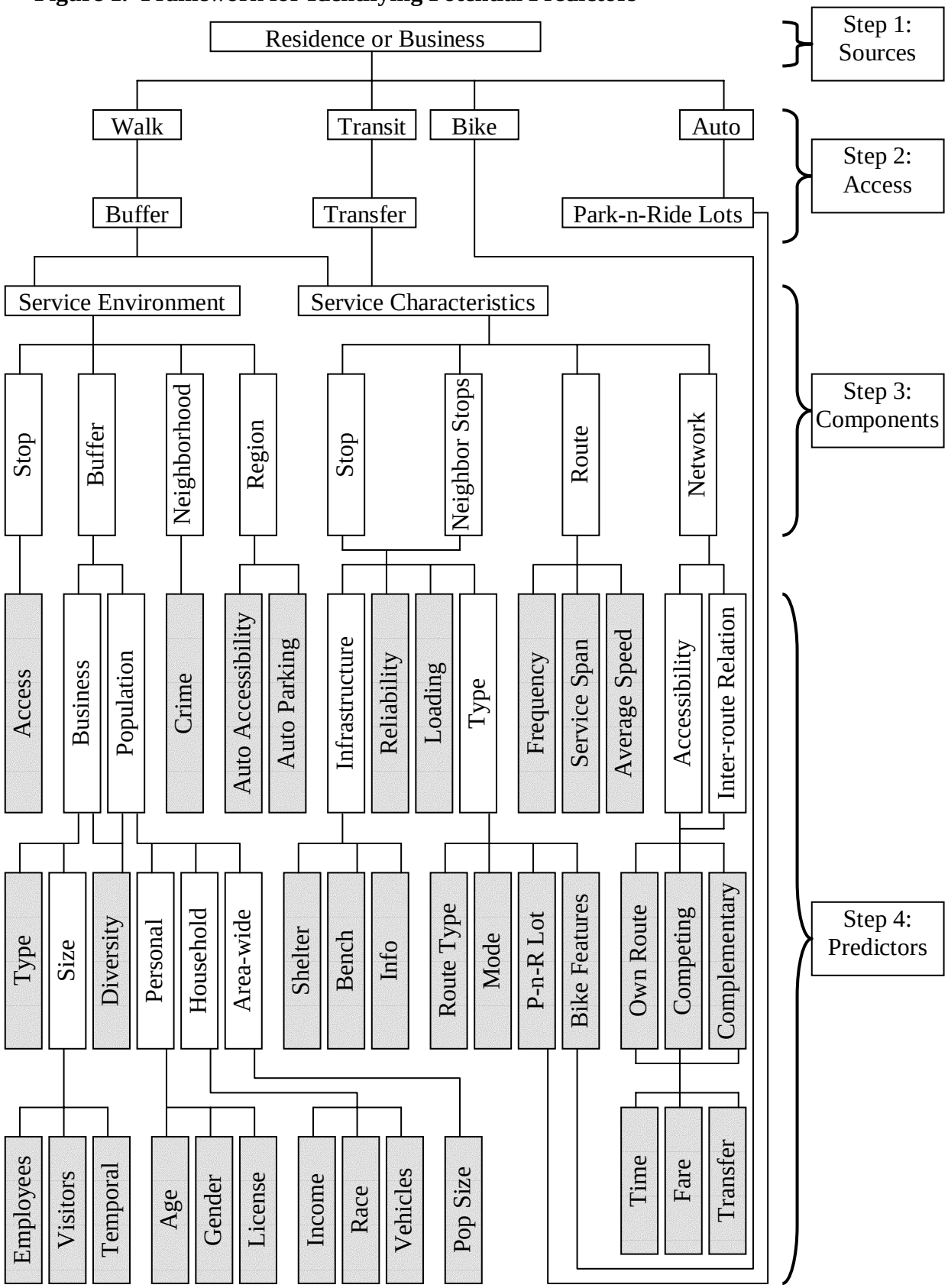
Figure 1 shows the framework used in determining the potential predictors of boarding at individual stops. This framework has four steps. The first step looks at the two main sources of boarding: residence and business. People board transit vehicles from home to go doing something (business); people board transit from businesses to go home or to another business.

The second step looks at the access modes before people board transit vehicles: walking, biking, transit, or auto. The geography from which people would be drawn to individual stops is likely to increase with these four access modes. For walking, it is the typical quarter-mile buffer. Little is known about the typical buffer size for biking. Data were not available to measure whether individual stops had designated areas for bike parking or whether the transit vehicles had bike racks. For transit, we are talking about transfers. For auto, we are talking about park-n-ride lots in terms of whether a park-n-ride lot is present or by the sizes of park-n-ride lots near individual stops.

The third step looks at the components of potential predictors for boarding for each of these access modes. The focus was on boarding with a walk access or transfer. There are two layers here. In the first layer, potential predictors are characteristics of either the service itself or the service environment. In the second layer, each of these two components may be characterized at varying scales. For the service environment, we have characteristics of the stops themselves, the buffers and neighborhoods around the stops, and the region in terms of the auto mode. For the service itself, we have service characteristics at the stop level, the segment level (neighboring stops), the route level, and the network level. Both the service environment and service characteristics influence boarding with a walk access. For boarding with a transfer, on the other hand, characteristics of the buffer areas are not likely to have major influence. The need to consider neighboring stops results from the fact that people do not always go to the nearest stop. When there are terrain differences, people may prefer walking down slope. When the first downstream stop is a time-point stop or has better amenities, people may prefer to use it, especially if they are early for the next transit vehicle.

The last step determines a set of potential predictors based on the various components from the third step. The boxes with shading show potential predictors in general, while the white boxes are just intermediate steps. This is not absolute, however. Because of the desire to fit this diagram on a single page, some of the specific predictors are not listed. For example, the access characteristics of individual stops may include many aspects, including linkage between stops and sidewalks, presence of sidewalks, street crossing facilities, crossing delays, etc.

Figure 1. Framework for Identifying Potential Predictors



It is important to note that every one of the key transit quality-of-service factors that are readily measurable is reflected in the above framework. According to the TCQSM, there are a total of 12 such factors as shown in Table 2.

Table 2. Quality of Service Factors and their Treatment

Factors	Framework Treatment
Service Coverage	Walking-distance buffers, park-n-ride lots, bike features for origins; nothing for destinations
Pedestrian Environment	Stop access and walking-distance buffer
Scheduling	TLOS, frequency, and service span
Amenities	Stop infrastructure
Transit Information	Stop infrastructure
Transfers	Through accessibility measure
Total Trip Time	Through accessibility measure
Cost	Through accessibility measure; nothing for auto costs
Safety and Security	Crime
Passenger Loads	Stop characteristics
Appearance and Comfort	Not readily measurable
Reliability	Stop characteristics

Planned Data Collection

KAI planed to collect data on a variety of variables based on its contract with the Department. Those potentially relevant to model development included outputs from the current version of TLOS and other raw data from various sources in Jacksonville. The lists below are based on an e-mail that KAI sent to CUTR on November 1, 2001.

Information from TLOS

The * indicates variables that can be calculated using either air or walk buffers.

Address/Buffer Level

- TLOS (percent minutes served), for time periods ranging from 15 minutes to 1 week*
- Number of routes serving address*
- Number of stops serving address*

Stop Level—Transit

- People located within x minutes downstream of a stop*
- People located within x minutes upstream of a stop*
- People located within buffers served by a stop*
- Jobs located within x minutes downstream of a stop*

- Jobs located within x minutes upstream of a stop*
- Jobs located within buffers served by a stop*
- Presence of shelter (yes/no)
- Presence of bench (yes/no)
- Presence of park-and-ride (yes/no)
- Number of routes serving stop
- Trolley service (yes/no)
- Skyway service (yes/no)

Stop Level—Pedestrian Environment

- Traffic signal in immediate vicinity (yes/no)
- Median type (none, raised median, center two-way left-turn lane)
- Number of lanes on street
- Pedestrian street-crossing delay
- TLOS pedestrian adjustment factor
- P.M. peak hour traffic volume
- Presence of continuous sidewalk in stop vicinity (yes/no)

TAZ Level

- Transit-auto travel time difference from one TAZ to all other TAZs

Other Information

KAI planed to collect data from seven other sources: voter registration data, motor-vehicle registration data, property appraiser data, business permit data, crime data, zoning data, and FSUTMS input data. KAI sent CUTR a list of potential variables from each source. The discussion is based on CUTR's understanding of what were available from these sources. This understanding was enhanced through several conversations with contact persons listed by KAI for each data source.

Data on voter registration, motor-vehicle registration, property appraisals, and business permits were address-specific. Crime data were aggregated into 54 sub-zones. The boundaries of these zones were described but not geo-coded. Zoning data were parcel-specific and geo-coded. FSUTMS input files included socio-demographic data for individual traffic-analysis zones (TAZs).

Voter Registration

Besides residential addresses, the only other mandatory piece of information relevant to this study was age. Other potentially useful pieces of information were not mandatory and contained missing data, including gender, race/ethnicity, and whether the voter had a Florida driver's license.

The voter registration file included voters registered some years back but had since moved away from the registered address. It excluded, on the other hand, persons under 18 years, non-citizens, or convicted felons. Persons under 18 were 26.3 percent of the population in 2000 (http://www.state.fl.us/edr/Population/census18.xls). The number of registered voters as of June 1, 2001 (445,233) was 56.9 percent of the estimated 2001 population (782,307) (http://www.state.fl.us/edr/Population/webpage10.txt and http://duvalelections.coj.net/votrstat.html).

Motor-Vehicle Registration

Data on the number of passenger vehicles and the number of other vehicles at the address level were available from the motor-vehicle registration. Based on a telephone conversation with the Florida DHSMV, the gender information determines whether a vehicle was registered to an individual person (female or male) or to a business (company). Some vehicles may have been registered at certain address but were being used by residents or businesses at some other addresses. Also vehicles that had been traded in would still be registered to the previous owners. It was unclear about the number of such vehicles.

Property Appraisal

Besides the property address, this database had information on property use, assessed value, and heated area. There were more than 150 individual property uses possible that could be separated into residential and non-residential. The non-residential ones could also be aggregated into a few categories. For non-residential uses, the size of the heated area could be used to indicate the size of the business, especially for retail businesses.

Occupational Licenses

Besides the business address, the only other potentially useful information was the description of the business (http://pawww.ci.jax.fl.us/pub/property/tables/lzone.htm).

Zoning Map Data

Two groups of information were available from the zoning map data as part of Jacksonville's GIS. One was the zoning of each parcel of land. The other group included categories of land use (http://cojmap.coj.net/htmlpage/landuse.htm). A total of 9 categories were established:

1. Residential
2. Commercial
3. Industrial
4. Recreation and Open Space
5. Public Buildings and Facilities
6. Historical Resources
7. Conservation
8. Agriculture
9. Wetlands

In addition to these categories, sub-categories were also available for residential, commercial and industrial uses:

Residential

- Rural residential: up to 2 dwelling units per acre
- Low density residential: up to 7 dwelling units per acre
- Medium density: up to 20 dwelling units per acre
- High density: 20 to 60 dwelling units per acre

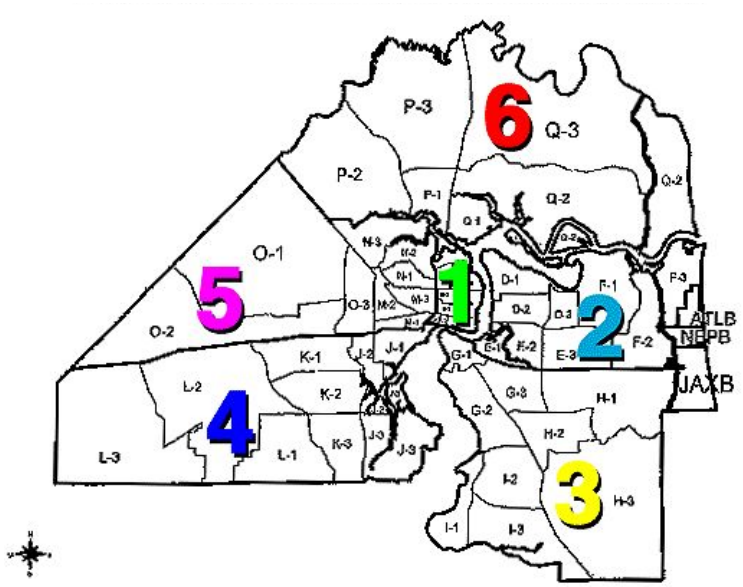
Commercial

- Residential-professional-institutional: mixed-use development
- Neighborhood commercial uses: within a 10-minute drive of the population
- Community/general commercial uses: serve large areas of the city
- Regional commercial: serve the city and outlying communities
- Central business district

Industrial

- Business park: low to moderate intensity office and industrial parks
- Light industrial: with fewer objectionable impacts
- Heavy industrial: with adverse physical and environmental impacts

Figure 2. Zones, Sectors, and Sub-sectors for Crime Data



Crime

The County Sheriff's Office maintained annual crime data by three levels of geography: zones (6), sectors (16), and sub-sectors (54). Figure 2 shows these geographies. The large numbers indicate the zones; the letters indicate sectors; and the smaller numbers indicate the sub-sectors.

The data for each sub-sector used the standard categories in the Uniform Crime Report: violent and property crimes. Violent crimes were murder, rape, other forcible sex crime, robbery, and aggravated assault. Property crimes included burglary, larceny, vehicle theft, and arson.

FSUTMS

The output files of FSUTMS contained travel time by auto between TAZ pairs. In addition, the input files contained population, population characteristics, and employment for individual TAZs. Specifically, two types of data were available for traffic analysis zones: trip production variables and trip attraction variables. Trip production variables measured characteristics of both residential population and hotel and motel population. Trip attraction variables measured the level of employment, school enrollment, and parking costs. These data were developed as part of FSUTMS. Trip production variables were contained in a file called ZDATA1, while trip attraction variables in ZDATA2.

Characteristics of Residential Population - Trip Production

- Single-family dwelling units
- Percent single-family dwelling units vacant
- Population in single-family dwelling units occupied by permanent residents
- Percent households having no vehicles in single-family dwelling units
- Percent households having one vehicle in single-family dwelling units
- Percent households having two or more vehicles in single-family dwelling units
- Multi-family dwelling units
- Percent multi-family dwelling units vacant
- Population in multi-family dwelling units occupied by permanent residents
- Percent households having no vehicles in multi-family dwelling units
- Percent households having one vehicle in multi-family dwelling units
- Percent households having two or more vehicles in multi-family dwelling units

Characteristics of Hotel/Motel Population - Trip Production

- Hotel-motel units
- Percent hotel-motel units occupied
- Persons in occupied hotel-motel units

Level of Employment - Trip Attraction

- Industrial employment by place of work
- Commercial employment by place of work
- Service employment by place of work

School Enrollment - Trip Attraction

- Total school enrollment

Comparing Potential Predictors with Planned Data Collection

The comparison was organized according to the framework in Figure 1. It started with boarding with bike or auto access, and then proceeded with the different components for walk and transit access.

Auto Access

The planned data collection had an indicator whether a park-n-ride lot was present at any stop. It was unclear, however, when a park-n-ride lot would be considered being present. Was it when a park-n-ride lot is within a quarter-mile buffer of a stop?

Bike Access

Bike access refers to bike features such as parking spaces and locking devices at stops or bike racks on transit vehicles. Few transit vehicles at JTA had bike racks. The planned data collection did not have any information on bike facilities either at stops or on buses.

Service Environment – Stop

The service environment at the stop level refers to the ease with which transit users can access individual stops. This access has at least two components: walking along the road where the stops are located and crossing it. The planned data collection included many relevant variables:

- Whether a traffic signal is present in the immediate vicinity of a stop (yes/no)
- Median type (none, raised median, center two-way left-turn lane)
- Number of lanes on the street
- Pedestrian street-crossing delay
- TLOS pedestrian adjustment factor
- P.M. peak hour traffic volume
- Presence of continuous sidewalk in the stop vicinity (yes/no)

The last variable in the above list intended to measure the ease with which transit users can walk along the street. All other variables help measure the difficulty with which transit users may experience in crossing the street. These variables would be mostly irrelevant to those users who do not need to cross the street.

Service Environment – Neighborhood

The neighborhood characteristics here focus on the safety and security of the area around individual stops. The planned data collection included the annual crime data from the Duval County Sheriff's Office. In order for the data to be useful, however, they needed to be disaggregated at the level of the 57-subsectors rather than at the level of 16 sectors.

Service Environment – Buffer

Business

The business portion of the service environment refers to the amount of opportunities available for transit users to participate in activities. Such opportunities are likely to vary with both the type of business and its size. The most relevant approach to categorizing businesses would be whether they are visitor-oriented or employee-oriented. Retail would be a good example of businesses that are visitor-oriented, while office space is a good example of businesses that are employee-oriented. For employee-oriented businesses, a measurement of the number of employees would be adequate because it is they who are the potential transit users. For visitor-oriented businesses, on the other hand, we would need a measurement of the size. For retail, square footage would be a good measure of its size. In addition to business types and sizes, also relevant is the temporal availability of the opportunities. When they are closed to either employees or visitors, businesses do not provide any opportunities for transit users to participate in activities.

The three databases that the planned data collection was going to use provided adequate data on business types at the address level.

The database from the Duval County Property Appraiser's Office provided adequate data on the size of businesses in terms of heated area.

None of these databases, however, contained information on the number of employees or the operating hours of the businesses.

The planned data collection did not include information on the number of jobs as used in calculating accessibility measures. The planned data source for the number of jobs was InfoUSA, which were measured at the address level.

Population

Population characteristics may be used at three levels: personal, household, and area-wide. Personal characteristics would include age, gender, and status of driver's license. Household characteristics would include income, race/ethnicity, vehicle ownership, housing, etc. Area-wide characteristics include population size and density, if the buffer areas differ in size.

Information was available on age and population size from the database on voter registration. It was noted that the registered age needed to be adjusted to reflect the fact that some registrations were done some time ago.

Vehicle ownership information was available from the motor-vehicle registration database.

There was no direct measure of household income from any of these databases. It was proposed to the property value of single-family houses to proxy household income. There was no similar proxy for households living in rental properties.

Diversity

The literature pointed to a three-dimension approach to describe the built environment: land-use density, land-use diversity, and pedestrian-friendly design (Cevero and Kockelman, 1997). We measured land-use density at the buffer level and pedestrian-friendly design at the stop level. We did not have any measure of land-use diversity at any level.

Service Environment – Region

This refers to the characteristics of the competing modes of transit. The auto mode was the focus. Two characteristics of the auto mode were used: accessibility and parking costs. Accessibility could be measured in two ways for any given boarding stop. One was to measure the average travel time by auto from the TAZ where this stop was located to every other downstream TAZ that had transit service, with the business activities in each TAZ as the weighting. The other was to measure the amount of business activities in all downstream TAZs that had transit service. This could be either cumulated within certain limits of auto travel time or weighted by auto travel time. Parking costs for any given stop refer to those at all potential downstream destinations served by transit, using the business activities at these destinations as the weighting to get an average measurement. The planned data collection had data on auto travel time and employment for TAZ pairs.

Service Characteristics – Stop

Infrastructure

Stop infrastructure refers to the amenities available at individual stops. The planned data collection included two aspects of stop infrastructure: shelter and bench presence.

Reliability

The planned data collection did not cover anything on reliability. To account for reliability at individual stops was difficult because regular stops did not have scheduled arrival and departure times that would have been needed to measure reliability by on-time performance. Since both time-point stops and others were going to be included, another approach was needed to measuring reliability at the stop level. One potential approach was to measure the day-to-day variation in the departure time of the same bus run at a given stop. The APC data did not have enough days for each run for this measurement.

Loading

The planned data collection did not cover loading at the stop level. Loading could potentially influence seat availability on the transit vehicles after boarding. Existing APC data would allow us to measure the average daily loading for individual stops. This average would need to be separately measured for the average weekday, Saturday, and Sunday, respectively.

Type

The planned data collection classified stops only in terms of mode. That was not enough. We needed to identify stops by the classifications proposed earlier.

TLOS

Data on TLOS were available at the stop level for the average weekday, Saturday, and Sunday, respectively. TLOS values account for several aspects of both the service environment and several service characteristics. These include frequency and daily service span. These also include the degree to which the build environment in the buffer area is pedestrian friendly. If TLOS did not account for it, the build environment would need to be a predictor at the buffer level in the service-environment component.

Neighboring Stops

When modeling transit boarding at the stop level, stop choice is part of the picture. Since we were not going to specifically consider stop choice, the consideration of the neighboring stops was intended to account for the potential competition these stops had on the stop in question.

The list of variables from the planned data collection did not include information on the neighboring stops. While the relevant information was available, the same variables for individual stops would need to be retrieved as new variables: one set for the immediately downstream stop and one set for the immediately upstream stop. If people were unlikely to walk to the upstream stop, we could do this just for the downstream stop.

Service Characteristics – Route

The list for the planned data collection did not include frequency or service span for individual routes. The data were readily available for the average weekdays, Saturday, and Sunday. One trick was that JTA typically adjusts its services on a quarterly basis. Another trick was that multiple routes served many of the stops. The planned variable list did not include the average operating speed by route either.

Service Characteristics – Network

Network Relationships

How well we could capture the relationships between individual stops and the rest of the transit network was critical to our success in modeling and predicting boarding at the stop level. The following lists several examples of such relationships that we needed to be able to measure:

- Two stops on the opposite sides of a road are likely to have very different relationships to the rest of the transit network. These two stops are similar otherwise; yet they frequently have very different boarding volumes.

- Some stops may serve both directions of the same route.
- Some stops may serve multiple routes that are either competing with or complementary to the route in question.
- Some stops may serve a single route but have near-by stops that serve other routes. Again these routes could be either competing with or complementary to the route in question.

The first example required the measurement of directional accessibility to population and businesses by transit.

We excluded the stops that serve both directions of a route.

The last two examples relate to the so-called competing routes and complementary routes. Some have treated such routes explicitly in a simultaneous equations system with the route in question. An alternative is to capture their presence in terms of service characteristics of the stop in question. More specifically, the presence of competing routes is not going to increase directional accessibility, while the presence of complementary routes is. In order to take this alternative approach, we needed to measure the incremental directional accessibility through the competing and complementary routes, respectively.

The only information on network relationships that the planned data collection included was the number of routes served by individual stops. We certainly needed to do more than that. First, we needed to determine our criteria in defining the following: 1) when a nearby route can potentially influence boarding at the stop in question; 2) whether a nearby route competes with the route in question; and 3) whether a nearby route complements the route in question. Second, we needed to measure stop activities at these stops as well as their service characteristics, including characteristics at the stop level, route level, and network level.

Travel Costs

It was equally important that we reflected the various components of travel cost by transit in our measurement of these network relationships. These included total trip time, fare, and the number of transfers. Because JTA used flat fare for a given population segment, reflecting fare in these measurements was not important in modeling JTA boarding. However, it was more important to the transferability of the resulting model because of differences in fare policy across transit systems in the state. The planned data collection included total trip time. We needed to expand that to include both fare and the number of transfers.

Additional Variables

Based on the above comparison, a set of additional variables was developed. These were grouped into those that required original data collection and those that needed to be derived from data already planned to be collected.

Derived from Existing Data

- Determine how businesses would be categorized
- Determine heated area for visitor-oriented businesses
- Property value of single-family housing
- Crime data for 57 sub-sectors
- Auto accessibility (average time or attraction)
- Reliability from APC data (weekday, Saturday, Sunday)
- Loading (weekday, Saturday, Sunday)
- Categorize stops into types and determine how to best characterize them
- Characteristics of neighboring stops
- Frequency and service span by route
- Average transit vehicle speed by route
- Determine criteria for competing and complementary routes
- Measure stop activities for competing and complementary routes
- Determine service characteristics of competing and complementary routes
- Measure directional accessibility by transit in terms of fare and number of transfers

Original Data Collection

- Bike features at stop and on buses
- Link individual stops with the 57 sub-sectors for crime data collection
- Determine employees for employee-oriented business
- Gender
- Race/ethnicity
- Income
- Parking costs
- Additional stop infrastructure

Recommended Final Set of Predictors

A teleconference was held on November 30, 2001 among CUTR, FDOT, and KAI to discuss the selection and measurement of variables. The recommended final set of variables was based on discussions during the teleconference along with several other steps:

- List of TLOS variables from KAI on 11/1/01;
- List of databases from KAI on 11/1/01;
- Further investigation of the databases by CUTR; and
- Analytical analysis of data needs from CUTR's working issue paper on 11/21/01.

These variables were grouped into three categories: dependent variables, predictors for the first-cut models, and additional predictors that might need to be added if the first-cut models turn out to be less than desirable.

Dependent Variables

- Annual average boarding: weekday daily, Saturday daily, Sunday daily, weekend day daily
- Annual average alighting: weekday daily, Saturday daily, Sunday daily, weekend day daily
- Average boarding since last service change: weekday daily, Saturday daily, Sunday daily, weekend day daily
- Average alighting since last service change: weekday daily, Saturday daily, Sunday daily, weekend day daily

Predictors for First-Cut Models

Stop Level—Transit

Some of these were in the original list from KAI. Some were moved from the Address/Buffer Level category in KAI's original list. The others were added from CUTR's analysis. Note that KAI planed to compute the number of downstream people or jobs as the cumulative among all buffer areas that were reachable from a given stop. These "reachable" buffer areas included stops on all routes serving that given stop. They did not, however, include stops on routes that did not directly serve that given stop but were nearby. Excluding the second group of reachable stops excluded competing and complementary routes that did not directly serve that given stop.

Original

- People located within x minutes downstream of a stop
- People located within x minutes upstream of a stop
- People located within buffers served by stop
- Jobs located within x minutes downstream of a stop
- Jobs located within x minutes upstream of a stop
- Jobs located within buffers served by stop
- Presence of shelter (yes/no)
- Presence of bench (yes/no)
- Presence of park-and-ride (yes/no)
- Number of routes serving stop
- Trolley service (yes/no)
- Skyway service (yes/no)

Moved from Address/Buffer Level

- Number of routes serving buffer area
- Number of stops serving buffer area
- Size of buffer area
- PMS (percent minutes served): weekday daily, Saturday daily, Sunday daily, weekend daily

Added

- Reliability: weekday daily, Saturday daily, Sunday daily, weekend day daily
- Loading: weekday daily, Saturday daily, Sunday daily, weekend day daily
- Stop types
- Average fare by mode
- People located within y transfers downstream of a stop: weekday daily, Saturday daily, Sunday daily, weekend day daily
- People located within y transfers upstream of a stop: weekday daily, Saturday daily, Sunday daily, weekend day daily
- Jobs located within y transfers downstream of a stop: weekday daily, Saturday daily, Sunday daily, weekend day daily
- Jobs located within y transfers downstream of a stop: weekday daily, Saturday daily, Sunday daily, weekend day daily
- Size of buffer areas downstream of a stop within
    - x minutes of transit time
    - y transfers
- Size of buffer areas upstream of a stop within
    - x minutes of transit time
    - y transfers

Stop Level—Pedestrian Environment

Every variable in this list was in the original list from KAI.

- Traffic signal in immediate vicinity (yes/no)
- Median type (none, raised median, center two-way left-turn lane)
- Number of lanes on street
- Pedestrian street crossing delay
- TLOS pedestrian adjustment factor
- P.M. peak hour traffic volume
- Presence of continuous sidewalk in stop vicinity (yes/no)

TAZ Level

- Transit-auto travel time difference from one TAZ to all other TAZs—as discussed previously, this could be weighted by the number of people/jobs in other TAZs.
- Auto travel time from one TAZ to all other TAZs—as discussed previously, this could be weighted by the number of people/jobs in other TAZs.
- Number of people within x minutes of auto travel time
- Number of jobs within x minutes of auto travel time

Other

The following were from new databases at the address level:

- Number of employees.
    - The Florida Department of Labor and Employment Security has address-level data on the number of employees.
- Number of licensed drivers.
    - The Florida Department of Motor Vehicles and Highway Safety has a database on licensed drivers, including their addresses.

The following were from KAI's original databases at the address level:

- Person age (from voter-registration database)
- Gender (from voter-registration database)
- Vehicles registered to residential addresses (from motor-vehicle registration)
    - Number of addresses with zero vehicles
    - Number of addresses with one vehicles
    - Number of addresses with two vehicles
    - Number of addresses with three+ vehicles
- Race/ethnicity (from voter-registration database)
- Number of persons (from voter-registration database)
- Square footage of heated area (from the property-appraisers database)
- Property values residential properties (property appraisers)
- Business types (to be derived from the business-permit database or the property-appraisers database). Preliminary types may include:
    - Residential – owner occupied
    - Residential – rental
    - Services
    - Retail
    - Industrial/wholesale
    - Government
    - Offices
    - Schools/churches
    - Recreational
    - Other
    - For a given type, businesses in the same buffer area will be aggregated by
        - Square footage of heated area
        - Number of employees.

The following were from KAI's original databases at other levels:

- Annual number of crimes by category for each of the 57 sub-sectors
- Relationship between stops and the 57 crime-data sub-sectors
- 2000 Census (Short-Form) – block level
    - Age
    - Gender
    - Race/ethnicity
    - Number of persons

- 2000 Census – block-group level
    - Average household income
    - Median household income
    - Percent of households with income under certain limit
    - We may use per capita property values as the proxy for per capita income. Specifically, per capita property values would be simply the ratio of the assessed property value and the number of residents.
    - One problem with this proxy is that a large portion of the variation in property values results from location differences.

Variables for Second-cut Models

These additional variables required changes to the currently proposed software structure for the new version of TLOS.

- Characteristics of neighboring stops upstream of a stop
- Characteristics of neighboring stops downstream of a stop
- Stop activities for competing routes
- Stop activities for complementary routes
- Stop characteristics (both transit and pedestrian) for competing routes
- Stop characteristics for complementary routes
- Number of jobs downstream of a stop via competing routes within
    - x minutes of transit time
    - y transfers
- Number of jobs downstream of a stop via complementary routes within
    - x minutes of transit time
    - y transfers

Limitations of Predictor Lists

The above list of predictors was largely for predicting stop boarding with walk or automobile access from residences. It was important for us to recognize that they had limitations if used to predict all boarding and alighting activities at the stop level. Several examples of such limitations are discussed below. This discussion can help us to overcome these limitations in future efforts of improving TLOS.

Stop Activities Related to Transfers

A JTA TDP under development indicated that about half of its riders transferred at least once on a daily basis. Suppose X was the daily number of linked trips. Assume that those riders who did transfer only transferred once. The daily number of unlinked trips would be about 1.5X. This means that transfers accounted for about one-third of the boarding and one-third of the alighting activities. Consequently it was critical that we did a reasonable job to predict the transfer component of boarding or alighting at individual stops. The key was to include the most important predictors for transfer activities.

It was important to note that characteristics of the buffer areas do not help predict transfer activities. What is critical is the relationship between any given stop and the rest of the transit network. Besides the number of jobs and people downstream or upstream of a given stop, the lists did not include specific predictors that described such a relationship. The cumulative number of jobs and people was unlikely to be adequate. At a minimum, we needed to separate the cumulative number of jobs and people via other routes at the given stop versus via other routes away from the given stop.

Stop Activities Related to Businesses

Given that transfer activities accounted for one-third of boarding and one-third of alighting activities, the other two-thirds were about evenly split between stop activities related to residences and businesses. Consequently it was equally critical that we did a reasonable job to predict the business-related component of stop activities.

For business-related alighting, it might be adequate to include predictors that described both the size and type of businesses. This may be accomplished with the number of employees and/or area-conditioned spaces within the buffer areas.

For business-related boarding, however, just the size and type of businesses may not be adequate. Additional predictors would be those that measure both the characteristics of the employees and the characteristics of the visitors at these businesses. A good job of describing business types may capture some of the employee characteristics but not those of the visitors.

MODEL FORMULATION ISSUES

We discuss several issues on model formulation and estimation. However, specific models will not be formulated because there is too much uncertainty on measurement.

Treatment of Competing and Complementary Routes

The issue is how competing and complementary routes should be modeled. Competing and complementary routes may be modeled explicitly with their stop activities or implicitly with the incremental accessibility they provide in addition to the route in question.

Some prefer explicitly modeling competing and complementary routes (Peng et al., 1997). In a simultaneous equations system for route-segment boarding, boarding on a subject route is modeled as a function of alighting on complementary routes and boarding on competing routes, while boarding on competing routes is modeled as a function of service on the subject route.

The explicit-modeling approach requires the determination of whether a route is competing or complementary. While it is relatively easy to determine whether two routes are independent to each other, it is far less clear how one may determine whether two routes are competing or complementary. In all cases, the determination is based on the degree of overlapping in buffer areas between two routes. Independent routes may be defined as those that do not overlap as

all. On the other hand, both a competing and a complementary route overlap with the subject route. The difference is in the degree. It becomes subjective to draw the line.

The implicit approach assumes that the downstream and upstream accessibility levels of a given stop to jobs and people fully capture the potential impact of competing and complementary routes on the given stop. The presence of complementary routes at a given stop is likely to increase its overall accessibility more than the presence of competing routes. Buffer areas on a complementary route are mostly net increases while buffer areas on a competing route overlap with those on the subject route.

In order for the implicit approach to be effective, however, it would need to account for two measurement factors. First, we need to consider all routes that serve the buffer area of the given stop in measuring both downstream and upstream accessibility. These routes include those that directly serve the given stop as well as those that do not. Second, we may need to separate accessibility provided by the competing and complementary routes serving the buffer area of the given stop from other components of downstream and upstream accessibility. Without this separation, we do not know whether a high level of accessibility results from a complementary route further downstream or from one that serve the buffer area of the given stop. A complementary route serving the buffer area of the given stop can directly influence activities at the given stop while a complementary route downstream cannot.

For the first cut of model development, it is recommended to use the implicit-modeling approach. The main advantage is that we do not need to define competing or complementary routes. A specific definition would require information on the spatial relationships among stops and routes. The current version or the planned new version of the TLOS software does not recognize such spatial relationships at this point. However, CUTR believes that it is important to consider all routes serving the buffer areas when measuring both downstream and upstream accessibility and to separately measure the accessibility of other routes serving the buffer area of the given stop.

Model

While many other factors can influence what is the most appropriate econometric model for a particular estimation problem, the issue is related to stop activities. The default approach is the standard linear regression model. If data are similar in other systems as LeeTran, the linear regression model is not likely to be appropriate for this research effort. The distributions of stop boarding, alighting, and total activities are extremely skewed toward the origin. Table 3 shows these distributions with data from the recent CUTR one-day ride check done for LeeTran. The first column lists the range of activities. These ranges become wider toward the bottom of the table. The other columns show the number of stops for each of the activity ranges. There are a total of 1,642 stops according to this ride check.

Table 3. Distribution of Stop Activities, LeeTran

Range	Boarding	Alighting	Total Activities
0	812	774	457
1-10	737	774	981
11-20	42	48	90
21-30	20	22	47
31-40	9	10	28
41-50	4	2	8
51-100	13	7	20
101-150	4	3	5
151+	1	2	6
Total	1,642	1,642	1,642

The alternative to the standard linear regression is to use count-data models (Greene, 2000). Distributions like these may mean that count-data based models, such as Poisson, negative binomial, and others, are more appropriate than the popular linear regression model. The Poisson regression model assumes that the mean and standard error are equal. That is frequently violated in real-life examples. The LeeTran data show that the mean and standard deviation are 3.15 and 10.91 for boarding and 3.15 and 10.73 for alighting. That is, the standard deviation is far greater than the mean. When this happens, the distribution is called over-dispersed. To allow for over-dispersion, one would need to use the called Negative-Binomial distribution. In addition, the high number of stops with no activities may require a model that can take into account this phenomenon. This is where zero-inflated Poisson, or negative binomial models, come in. The best strategy at this point would be to have the data ready so that these alternatives can be tested later.

Correlation between Boarding and Alighting

This refers to the high correlation between boarding at a stop and alighting at the stop on the opposite side of the route. In case the Department decides to model alighting as well, we need to figure out an approach to dealing with this high correlation. One approach would be to treat boarding and alighting separately. This single-equation approach fails to take advantage of the correlation. While the resulting coefficients would not be biased, they are not efficient. That is, the standard errors of the coefficients would be over-estimated. An alternative approach would be to use the so-called seemingly unrelated regression equations (SURE) that takes advantage of this correlation (Winkelmann, 2000). The SURE results would be identical to those of single-equation models if the same set of predictors is used for all equations or if boarding and alighting are not correlated. It is reasonable to expect at this point that boarding and alighting should have different sets of predictors. Accessibility to jobs and people downstream of a stop is more relevant to boarding at that stop, for example, than to alighting at the opposite stop. The best strategy at this point would be to have the data ready so that both can be tested later.

Endogenous Problem

There are several groups of potential predictors that could result in endogenous problems if included. An endogenous problem refers to the correlation between one or more predictors and the error term in a stop-activity model. Without appropriate treatment, the coefficients of these potential predictors would be biased. These predictors include supply variables, stop infrastructure variables, and land use variables.

Service Supply

The issue is whether the model should be estimated as part of a simultaneous-equations system between stop activities and service supply, or as a single equation for stop activities. There are several considerations involved.

- Whether supply and stop activities are truly determined simultaneously in reality. In the Portland efforts on modeling boarding at time-point stops, for example, the researchers decided to use the single equation approach because they believed that boarding is stop-specific while frequency is route specific. In the case of this research, entering the supply variables in a stop-activity model does not appear to cause the endogenous problem. As a result, estimating a stop-activity model would not result in bias.
- Whether the prediction of stop activities should be conditional on supply levels. If the prediction need not be conditional on supply levels, a reduced-form stop-activity model would be estimated without the supply variables entered. This was the approach taken in one national study of stop boarding of light rail and commuter rail (PBQD et al., 1996b). An alternative reason for not including supply variables would be that they do not vary across observations. This was the approach taken in a study of BART station activities (PBQD et al., 1996a).
- Whether a supply equation is needed for prediction later. In ITSUP, for example, the simultaneous-equations approach was taken, and both the supply and boarding equations were used in prediction (FDOT, 2001). In the case of this research, on the other hand, only stop activities will be predicted.

It is recommended to use single equations for stop activities and to include TLOS values in the stop-activity equations. These are based on several assumptions:

- We do not have an endogenous problem between supply and stop activities
- We do not need to predict supply
- The prediction of stop activities will be conditional on supply levels.

Stop Amenities

Unlike including supply variables in stop-activity models, using stop amenities in such models is likely to cause an endogenous problem. The reason is this: Many stop amenities are put in place by transit agencies because of high stop activities. The consequence is biased estimates of the predictor coefficients. There are three general approaches to dealing with this potential endogenous problem.

- By estimating models without stop amenities, we get reduced-form models. While such models do not explicit include stop amenities, they effects are implicitly accounted for. Casual evidence seems to suggest that stop amenities are installed to serve existing high volumes of boarding or alighting rather than to attract new customers. Also, the presence of stop amenities may have a larger effect on stop selection than on attracting new riders.

- Another alternative would be to include stop amenities without specifically treating the potential endogenous problem. This approach is likely to result in serious biases in estimated coefficients.

- The third approach would be to include stop amenities and to explicitly account for the resulting endogenous problem. This would be ideal if we can find a way to do it.

Two other modeling choices may complicate our efforts to take this approach. One is whether we model both boarding and alighting in a seemingly unrelated equations system. The second modeling choice is whether we use econometric models that can account for the discrete nature of boarding or alighting and the large number of stops without any activities. The complication results from the fact that commercially available software tools do not appear to be equipped to estimate these relatively complex models yet.

It is recommended to take the reduced-form approach and test the third approach when the chosen software allows such a test.

Land Use

People have argued that land use characteristics, especially density, are jointly determined with transit usage. One reason is that people are likely to select to live in high-density areas if they use, or plan to use, transit. The approaches are similar to how one may deal with stop amenities: 1) Exclude land use; 2) Include land use but ignore the potential endogenous problem; and 3) Include land use and account for the potential endogenous problem. The first approach is to estimate reduced-form models in terms of land use. This approach is unsatisfactory because of the potential importance of land use. Compared to stop amenities, the seriousness of the endogenous problem resulting from land use, if exists, is likely to be minor. To the extent that this assessment is correct, the second approach is a reasonable alternative. It is recommended to ignore the potential endogenous problem from using land use as predictors.

Econometric Software

In order to explore the different econometric models for count data, the project will need to use a piece of econometric software that can accommodate that. Based on our preliminary examination, the best candidate seems to be LIMDEP. However, we still need to find a software package that could estimate a simultaneous equations system with a count-data model and a binary choice model if we want to include stop amenities as predictors. In addition, there appears no commercial software available that allows the user to directly estimate a model

system with both count data and seemingly unrelated regression. Furthermore, there are no software tools for directly estimating such models with an endogenous selection problem. One possible way out of these issues would be to use an econometric software package that allows the user to indirectly estimate a model system through programming.

PART II: EMPIRICAL WORK

This part documents the data available and used, and the selection of a final model.

ESTIMATION DATA

KAI, in cooperation with URS, Inc., provided the data for model estimation. The data made available by KAI differ from the specification described in Part I. This difference is discussed first. The data used in model estimation also differ from what were available. This difference is discussed next.

Data Available

Table 4 shows the variables, their brief descriptions, and data sources for which data are provided by KAI. These data deviate from what were described earlier. The following are just a few examples.

One deviation is on the level of aggregation of socio-demographics. First, the FDOT wanted to use address-level data for all social demographics. However, all usage data available on social demographics are derived from aggregate data. See the next section for discussion on usable data. Second, crime data are available at three geographic levels with the most detailed level dividing the Jacksonville area into 54 sections. However, the crime data provided are at a level with only 16 geographic sections.

A second deviation is on the absence of a number of variables on the recommended final set of predictors described earlier. Examples of such variables include auto accessibility, transit reliability at the stop level, stop categories, route types, special generators, square footage of commercial heated area at the address level, and residential property values as a substitute of income at the address level. Park-n-Ride lots are examples of special generators. The absence of Park-n-Ride lots prevents the consideration of their role in boarding at the stop level.

A third deviation is on the unit of analysis. One possible unit of analysis would be stops that are spatially unique. Two two-direction routes intersecting at an intersection would have four spatially unique stops in this sense. For the purpose of reducing processing time as part of TLOS Version 4, FDOT decided to aggregate these four stops at an intersection into a so-called TLOS stop. This aggregated unit of analysis is also used in the data provided by KAI.

A fourth deviation is on the aggregation of directions for boarding volumes at any stop locations. This deviation is really a by product of the third deviation. The use of TLOS stops makes it impossible to separate boarding by direction because a TLOS stop at an intersection where two two-direction routes intersect, for example, includes all four actual stops.

Table 4. Data Available

Variable	Description	Source
NODEID	stop id	TLOS
routes	routes serving stop	TLOS
bufroutes	routes serving stop's buffer area	TLOS
bufstops	stops within buffer area	TLOS
bufarea	stop buffered area	TLOS
area_pop	people at stop by percent area	Census 2000
area_job	jobs at stop by percent area	TAZ
road_pop	people at stop by percent road	Census 2000
road_job	jobs at stop by percent road	TAZ
pdown0xfer	people downstream (no Xfer) weekday/Saturday	Census 2000
jdown0xfer	jobs downstream (no Xfer) weekday/Saturday	FDLES
pdown1xfer	people downstream (1 Xfer) weekday/Saturday	Census 2000
jdown1xfer	jobs downstream (1 Xfer) weekday/Saturday	FDLES
pup0xfer	people upstream (no Xfer) weekday/Saturday	Census 2000
jup0xfer	jobs upstream (no Xfer) weekday/Saturday	FDLES
pup1xfer	people upstream (1 Xfer) weekday/Saturday	TLOS
jup1xfer	jobs upstream (1 Xfer) weekday/Saturday	Census 2000
WKDYTLOS	PPMS Weekday	TLOS
SATTLOS	PPMS Saturday	TLOS
SUNTLOS	PPMS Sunday	TLOS
arriv_wd	Avg number of bus arrivals weekday	TLOS
arriv_sa	Avg number of bus arrivals Saturday	TLOS
arriv_su	Avg number of bus arrivals Sunday	TLOS
trolley	trolley stop	TLOS
skyway	skyway stop	TLOS
Shelter	shelter	Field Survey
Bench	bench	Field Survey
signal	traffic signal	Field Survey
lanes	number of lanes on street	Field Survey
pfactor	TLOS pedestrian adjustment factor	Kittelson
board_wd	avg number persons entering bus at stop	JTA
alight_wd	avg number of eprsons exiting bus at stop	JTA
board_sat	avg number persons entering bus at stop	JTA
alight_sat	avg number of eprsons exiting bus at stop	JTA
board_sun	avg number persons entering bus at stop	JTA
alight_sun	avg number of eprsons exiting bus at stop	JTA
ageunder18	Persons aged 18 or younger	Census 2000
age18_64	Persons aged 18 to64	Census 2000
age65over	Persons aged 65 or older	Census 2000
male	Male persons	Census 2000
female	Female persons	Census 2000
white	Persons of white race	Census 2000
nonwhite	Persons of non-white race	Census 2000
hispanic	Persons Hispanic	Census 2000
medinc	Median household income	Census 2000
povrty	Share of households under poverty	Census 2000
veh0	Addresses with 0 vehicles	DMV
veh1	Addresses with 1 vehicles	DMV
veh2	Addresses with 2 vehicles	DMV
veh3plus	Addresses with 3+ vehicles	DMV
voters	Registered voters	Voter Registration
avghharea	Average residential HH heated area (sq. ft.)	Property Appraiser
avghhvalue	Average residential HH value($)	Property Appraiser
avgbusarea	Average business heated area (sq. ft.)	Property Appraiser
avgbusvalue	Average business value ($)	Property Appraiser
businesses	Number of businesses by type	InfoUSA
drivers	licensed drivers	License data
crimes_violent	Number of violent crimes in stop's sector	JSO
crimes_property	Number of property crimes in stop's sector	JSO

Data Used

Many variables provided by KAI are not used in estimating the final model. These fall into three groups. The first group includes variables measured at the address level but its quality is low. These include employment data from InfoUSA, voter data, number of drivers derived from state license data, and number of addresses with different levels of vehicle ownership. For some of these variables, data are missing for a large number of stops. For some of the others, data are simply wrong. The number of addresses with zero vehicles, for example, is negative for many catchment areas.

The second group includes dependent variables that require independent variables that are not available. Sunday boarding was not considered because data on accessibility measures are unavailable. Saturday boarding was not considered because data on the components of TLOS are unavailable.

The third group includes alighting variables. Alighting was not considered because it is highly correlated with boarding. In fact, the correlation coefficient between alighting and boarding is 0.92 for weekdays, 0.95 for Saturdays, and 0.96 for Sundays. The best prediction that one can do for alighting is likely to be boarding.

ESTIMATION

Steps

CUTR focused on finding the best model possible with the usable data described above through several steps. These steps are largely sequential but are also interactive to some degree.

The first step was to define what constitutes a model to be further considered in this process of searching for the best model. Two aspects are involved. One aspect was to categorize all potential explanatory variables into several groups: 1) socio-demographics in the catchment area around a TLOS stop; 2) service supply; 3) street environment; 4) accessibility; 5) interaction with other modes; and 6) competition from other TLOS stops. The other aspect was to define the expected signs of coefficients for these variables. Models that do not meet these expectations were not further considered.

The second step was to consider a variety of alternatives models. These alternatives represent variations in several dimensions. These dimensions include the statistical model, socio-demographics, components of TLOS, components of accessibility measures, number of transfers in measuring accessibility, time limit on accessibility measures, and stratification of TLOS stops. Each of these dimensions is briefly described below:

- Alternative statistical models represent differences in both the dependent variable and the statistical model used. The dependent variable uses one of two forms: 1) estimated boarding using APC data (or APC boarding); and 2) estimated boarding rounded to the nearest integer (or rounded boarding). The statistical model also takes one of two forms: 1) the standard linear regression model; and 2) regression models for count data. The basic count model is

the standard Poisson regression model. Other count models are also available. These alternatives are considered because the most appropriate statistical model depends on the nature of the dependent variable.

- There are many elements related to the socio-demographics of TLOS stop catchment areas, including employ, population, and their characteristics.

- Stop amenities do attract ridership to some degree and should be included if their effects can be correctly estimated. Directly including variables indicating the presence of these amenities will likely over-estimate their effect on boarding. These amenities frequently are provided because of higher boarding volumes.

- For both population and employment, accessibility is measured separately for upstream and downstream directions. The options are either to keep the directions separated or to combine them. This is done largely to reduce the number of variables and to reduce ambiguity in coefficient signs.

- The amount of population or employment reachable from a subject TLOS stop depends on both the maximum travel time allowed and the number of transfers allowed.

- The options on TLOS are either to use the total week day TLOS or to break it into two components: the component based on one-minute walk buffers around a TLOS stop and the rest of TLOS. The expectation is that the two components may have differential effects on boarding.

- Accessibility measures are computed for several time limits, including 30 minutes, 40 minutes, and 60 minutes. There is no theoretically correct limit to use in defining accessibility measures. The shorter limits were used to see if they may perform better than the 60-minute limit.

- TLOS stops were stratified by the number of routes serving the catchment area of a subject TLOS stop. Since most of the available variables represent potential determinants of direct boarding rather than transfer boarding, these variables may do a better job of predicting total boarding at TLOS stops whose catchment areas are served by a single route, for example.

- The Public Transit Office also wanted to segment the stop population by boarding ranges. The idea was that if one cannot develop a single satisfactory model for all stops, can one develop satisfactory models for stops with different ranges of ridership?

The third step of the searching process for the best model was to evaluate the alternative models in terms of their fit to the data and their prediction accuracy. Their fit to the data is measured with the improvement in log likelihood computed as follows:

$$\rho^2 = 1 - \frac{\text{Log} - \text{Likelihood}}{\text{Restricted Log} - \text{Likelihood}}$$

where log-likelihood is the value when a full model is estimated while restricted log-likelihood is the value when the model with only a constant is estimated. Their prediction accuracy is measured with the root mean square error (RMSE) computed as follows:

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{\left(\text{Predicted}_i - \text{Actual}_i\right)^2}{n}}$$

where the summation is cross individual TLOS stops in the entire sample.

Model Selection

There would have been a large number of comparisons had the evaluation been done on each possible pair of alternatives. Instead, comparisons on log-likelihood and prediction accuracy were made in terms of time limits on accessibility measures and stratification of TLOS stops when the other components of model specifications were already established. These components are:

1. It was established early on that the Poisson regression model with rounded boarding is preferred than the standard linear regression model. One significant problem with the linear regression model is that the estimated coefficients of several of the independent variables have the wrong sign. These include TLOS, accessibility to jobs, and the number of other TLOS stops in the catchment area. In addition, several variables are statistically insignificant, including the median household income in the catchment area and the pedestrian factor. By using the Poisson regression model, on the other hand, these two problems of the linear regression models have disappeared. In addition, the Poisson regression model outperforms in terms of prediction accuracy. The average RMSE from the Poisson regression model is almost half of that from the standard linear regression model. Other count models were also considered. These models did not lead to conversion in estimation, however.

2. The following set of socio-demographics has a robust effect on weekday boarding:

- Median household income (000s)
- Number of jobs in catchment area by percent road
- Zero-vehicle households in catchment area
- Share of persons under 18
- Share of persons 18-64
- Share of persons female
- Share of persons Hispanic
- Share of persons White

3. Stop amenities were excluded from all models considered for two reasons. One is that the amenities are not well defined for TLOS stops because a single TLOS stop can contain multiple regular stops. A presence of a bench at a TLOS stop, for example, can mean the presence of a bench only at one regular stop or at all regular stops included in the TLOS stop. The other reason is that the data do not one to correctly estimate the true effect of these amenities.

4. It was established early on that combining the upstream and downstream accessibility measures leads to more stable results, i.e., less fluctuation in coefficient signs and the statistical significance of coefficients.

5. It was established early on that accessibility measures defined with up to one transfer do not do better than those defined with no transfers allowed.

6. It was established that the two TLOS components perform better than total TLOS.

7. It was empirically established that segmenting the stop population by ridership ranges does not show improvements in the models. More important, such segmented models are useless for forecasting ridership at the stop level, which would require knowing the ridership level before forecasting is carried out.

Some of the models estimated with these considerations are presented in the Appendix. These models are just some of those estimated during a work session between CUTR and the FDOT Project Manager on February 16, 2004.

With the basic specification of these seven elements of the model, it was further tested in terms of time limit on accessibility measurement and stratification of TLOS stops. It was clear from the testing that the lower time limits (30 minutes and 40 minutes) do not improve the model. Furthermore, stratifying TLOS stops by the number of routes serving individual catchment areas did not improve the model either.

Final Model

With the usable data, carrying out the three steps discussed above led to the following final model of weekday total boarding at individual TLOS stops in Table 4. The model is satisfactory in three important ways. First, as expected, all six categories of variables influence ridership at the stop level for average weekdays. In addition to the traditional variables measuring the characteristics of catchment areas, TLOS measures, the pedestrian environment, accessibility measures to population and employment, interactions with other modes, and competition from other stops in catchment areas all play a statistically significant role in stop patronage for average weekdays. Second, all of these variables have expected coefficient signs. Third, these variables as a group fit the data well. The presence of these variables improves the log likelihood value by 54 percent as shown at the bottom of Table 4.

Table 4. Poisson Regression of Weekday Total Boarding at TLOS Stops

Independent variables	Coeff.	t-ratio	Mean
Constant	0.2152	3.42	
Median household income (000s) in catchment area	-0.0045	-5.63	35.157
Jobs in catchment area by road	0.0001	26.28	390.503
0-vehicle households in catchment area	0.0028	13.95	22.059
Share of persons under 18 (0 to 1) in catchment area	-0.6122	-6.60	0.240
Share of persons 18-64 (0 to 1) in catchment area	0.4978	12.69	0.606
Share of persons female (0 to 1) in catchment area	0.5843	8.25	0.510
Share of persons Hispanic (0 to 1) in catchment area	5.3349	22.42	0.033
Share of persons White (0 to 1) in catchment area	-0.5075	-14.79	0.507
TLOS within one-minute walking (0 to 100)	0.0458	70.12	20.129
TLOS within two-five minutes walking (0 to 100)	0.0195	26.19	17.370
Pedestrian factor (0 to 1)	0.1203	7.12	0.658
Persons up and downstream without transfer (000s) in 1 hr	0.0026	11.78	26.313
Jobs up and downstream without transfer (000s) in 1 hr	0.0027	8.25	50.264
Including a trolley stop (1 if present; 0 otherwise)	0.2176	5.25	0.003
Number of other TLOS stops in catchment area	-0.0783	-26.18	4.157
Log likelihood	-18,072		
Restricted log likelihood	-39,063		
Goodness of fit	0.54		
Sample	2,568		

SUMMARY

This project developed an Issue Paper in December 2001 that described the research design in terms of previous work, a framework, data needs, model estimation issues. In addition, the Issue Paper outlined alternative approaches addressing the various issues encountered in formulating ridership models at the stop level. This is reflected in Part I of this report.

Part II of this report presents the development of a final model, using data provided by Kittelson and Associates, Inc. (KAI) under a separate contract with the Department. The KAI data deviated from the specifications in the Issue Paper in a number of ways. Instead of address-level data for many variables, all usable data for social demographics were derived from aggregated data. Many variables on the specified list were missing, including auto accessibility, transit reliability at the stop level, stop categories, route types, special generators (e.g., park-n-ride lots), square footage of commercial heated area at the address level, and residential property values as a substitute of income at the address level. Instead of using spatially unique stops as the unit of analysis, the FAI data used TLOS stops, which aggregate all spatially unique stops around intersections. Finally, instead of considering boarding by direction at spatially unique stops, the use of TLOS stops makes it impossible to separate boarding by direction.

CUTR developed a final model of average weekday boarding for TLOS stops following a comprehensive statistical analysis of the usable data provided by KAI according to the conceptual considerations in the Issue Paper. The model is satisfactory in three important ways. First, in addition to the traditional characteristics of catchment areas, the TLOS indicator, the pedestrian environment, accessibility measures to population and employment, interactions with other modes, and competition from other stops in catchment areas all play a statistically significant role in average weekday boarding. Second, all of these variables have expected coefficient signs. Third, these variables as a group fit the data well. The presence of these variables improves the log likelihood value by 54 percent.

ACKNOWLEDGMENT

The Florida Department of Transportation funded the research through the National Center for Transit Research. Jon Ausman was the Project Manager for the Florida Department of Transportation. Paul Ryus of Kittelson and Associates, Inc. and Marc Cooper of URS, Inc. provided the data.

REFERENCES

Alperovich G, Kemp MA, and Goodman KM (1977) *An Econometric Model of Bus Transit Demand and Supply*. Working Paper 5032-1-4. The Urban Institute, Washington, D.C.

Atherton TJ and Ben-Akiva ME (1976) Transferability and updating of disaggregate travel demand models. Transportation Research Record 610: 12-18.

Ausman, Jon (2000a). Predictability of Transit Use, e-mail to the Principal Investigator on May 12, 2000.

Ausman, Jon (2000b). Responses to Queries, e-mail to the Principal Investigator on May 15, 2000.

Ausman, Jon (2001). Verbal conversation with the Principal Investigator.

Boyle DK (1998) *Passenger Counting Technologies and Procedures*. TCRP Synthesis of Transit Practice 29, Transportation Research Board, Washington, D.C.

Center for Urban Transportation Research (2001). Spreadsheet file that contains the count data from a ride check for LeeTran.

Cervero, R; Kockelman, K (1997) Travel Demand and the 3Ds: Density, Diversity, and Design. Transportation Research-Part D 2: 199-219.

Cherwony W and Polin L (1977) Forecasting patronage on new transit routes. *Traffic Quarterly* 31: 287-295.

Florida Department of Transportation (2001a), *State Part & Ride Lot Program Planning Manual*.

Florida Department of Transportation (2001b), *Integrated Transit Demand and Supply Software*, http://www.dot.state.fl.us/Transit/itdsup/itdsup.htm.

Greene WH (2000) *Econometric Analysis*. Fourth Ed., Prentice Hall, Upper Saddle River, N.J.

Hartgen D. and Horner MW (1997) *A Route-Level Transit Ridership Forecasting Model for the Lane Transit District: Eugene, Oregon*. Report No. 170, Center for Interdisciplinary Transportation Studies, Charlotte, North Carolina.

Horowitz AJ (1984) Simplifications for single-route transit ridership forecasting models. *Transportation* 12: 261-275.

Horowitz AJ and Metzger DN (1985) Implementation of service area concepts in single-route ridership forecasting. *Transportation Research Record* 1037: 31-39.

Kemp MA (1981) *A Simultaneous Equations Analysis of Route Demand, Supply, and Its Application to the San Diego Bus System*. Report No. 1470-2, The Urban Institute, Washington, D.C.

Kikuchi S. and Miljkovic D (2001) Use of fuzzy inference for modeling prediction of transit ridership at individual stops. *Transportation Research Record* 1774: 25-35.

Kimpel TJ, Strathman JG, and Dueker KJ (2000) *Time Point-Level Analysis of Passenger Demand and Transit Service Reliability*. Center for Urban Studies, College of Urban and Public Affairs, Portland State University, Portland, Oregon.

Kittelson & Associates, Inc., and URS, Inc. (2000). TLOS Software Users Guide, Version 3.0.

Koppelman FS and Wilmot CG (1982) Transferability analysis of disaggregate choice models. *Transportation Research Record* 895: 18-24.

Menhard HR and Ruprecht GF (1983) Review of route-level ridership prediction techniques. *Transportation Research Record* 936: 22-24.

Multisystems, Inc. (1982) *Route-Level Demand Models: A Review.* DOT-1-82-6, Urban Mass Transportation Administration, US Department of Transportation, Washington, D.C.

Parsons Brinckerhoff Quade & Douglas, Inc., Robert Cervero, Howard/Stein-Hudson Associates, Inc., and Jeffrey Zupan (1996a), *Transit and Urban Form*, TCRP Report 16, Volume 1, Part II, *Commuter and Light Rail Transit Corridors: the Land Use Connection*, Appendix A.

Parsons Brinckerhoff Quade & Douglas, Inc., Robert Cervero, Howard/Stein-Hudson Associates, Inc., and Jeffrey Zupan (1996b), *Transit and Urban Form*, TCRP Report 16,

Volume 1, Part II, *Commuter and Light Rail Transit Corridors: the Land Use Connection*, Chapter 3.

Pendyala RM and Ubaka I (2000) Development of short-term operational planning model for transit service analysis. *Transportation Research Record* 1735: 43-50.

Peng Z and Dueker JK (1995) Spatial data integration in route-level transit demand modeling. *Journal of the Urban and Regional Information Systems Association* 7: 26-37.

Peng Z, Dueker KJ, Strathman J, and Hopper J (1997) A simultaneous route-level transit patronage mode: demand, supply, and inter-route relationship. *Transportation* 24: 159-181.

Quandt, Richard E. (1970), *The Demand for Travel: Theory and Measurement*. Lexington, Mass., Heath Lexington Books.

Ryus P, Ausman J, Teaf D, Cooper M, and Knoblauch M (2000) Development of Florida's transit level-of-service indicator. *Transportation Research Record* 1731: 123-129.

Stopher PR (1992) Development of a route level patronage forecasting method. *Transportation* 19: 201-220.

Stopher PR and Mulhall S (1992) *Route Level Patronage Forecasting Methods: A Survey of Transit Operators*. Presented at the 71st Annual Meeting, Transportation Research Board, Washington, D.C.

Strathman JG, Dueker KJ, and Z Peng (1997) *Issues in the Design of a Stop-Level Transit Patronage Model*. Project Report PR 102. Center for Urban Studies, College of Urban and Public Affairs, Portland State University, Portland, Oregon.

Thompson, Gregory L. (1997) Achieving Suburban Transit Potential. *Transportation Research Record* 1571:151-162.

Transportation Research Board (2003) *Transit Capacity and Quality of Service Manual*, 2nd Ed., NCTR Report 100. National Research Council, Washington, D.C.

Winkelmann R (2000) Seemingly unrelated negative binomial regression. *Oxford Bulletin of Economics and Statistics* 62: 553-560.

APPENDIX: SELECTED TEST RUNS

The appendix shows the results of selected test runs jointly conducted between CUTR and the FDOT Project Manager on February 16, 2004. These results were regenerated on December 14, 2004 using the original specifications. The results are in the original form of the software used for model estimation (LIMDEP).

The test runs used a core set of independent variables that are common across them, including:

MEDINC	=	Median household income in catchment area
ROAD_JOB	=	Number of jobs by percent road in catchment area
VEH0HH	=	Number of households without vehicles in catchment area
AGE18_S	=	Share of persons under 18 in catchment area
AGEO_S	=	Share of persons from 18 to 64 in catchment area
FEM_S	=	Share females in catchment area
HISP_S	=	Share of Hispanic population in catchment area
WHITE_S	=	Share of White population in catchment area
PFACTOR	=	Pedestrian factor
P0XSAT	=	Sum of up and downstream population accessibility in 1 hour
J0XSAT	=	Sum of up and downstream jobs accessibility in 1 hour
O_STOPS	=	Number of other TLOS stops in catchment area

Note that this core set of independent variables does not include the dummy variable used in the final report that indicates the presence of a trolley stop at the subject TLOS stop.

The test runs vary in the day of week modeled, in the range of boarding used, in the measure of service supply, and the stratification of TLOS stops. Both Saturday and weekdays were considered. The different ranges of boarding volumes considered include the full range, zero boarding, no more than 10 boarding, no more than 30 boarding, and between 10 and 30 boarding. The measure of service supply includes daily five-minute TLOS for Saturdays, frequency, and the separation of daily one-minute TLOS (dly_tlos) and daily two- to five-minute TLOS (ndy_tlos) for weekdays. Arriv_wd is used for weekday frequency, and arriv_sa for Saturday frequency. The stratification of TLOS stops include stops served by any number of routes, stops served by a single route, and stops served by two routes.

After a brief description of the data and script used for conducting these test runs with LIMDEP, the LIMDEP results for individual test runs are shown first for Saturday and then for Weekday.

LIMDEP FILES

The attached CD contains a LIMDEP script file (testruns.lim) and one LIMDEP data file (testruns). The script file contains the LIMDEP commands for the test runs shown in this Appendix. If you have LIMDEP installed on your computer, follow the following steps to redo these test runs:

- Once you are in LIMDEP, enter 10000000 at: tools/options/projects/data cells.
- Insert the attached CD in your CD drive.
- Open the script file.
- Specify the CD drive in the LOAD command in the script file. The default is f:/
- Highlight the LOAD command and run it to have the data available for model estimation.
- Once the data are in, you are ready to create new variables and to estimate models.

OTHER FILES

The attached CD also contains the Word file of this report (Final report-BC137-31.doc), a summary of this report (BC137_31-summary.doc), and an Excel file (testruns.xls) of the same data contained in the LIMDEP data file described above.

SATURDAY RUNS

Five test runs are shown for Saturday boarding. They vary in the range for boarding and the measure of service supply. Stops may serve any number of routes. These runs are:

1. Full range of boarding, daily five-minute TLOS, stops served by any number of routes
2. Full range of boarding, frequency, stops served by any number of routes
3. No more than 10 boarding, daily five-minute TLOS, stops served by any number of routes
4. No more than 30 boarding, daily five-minute TLOS, stops served by any number of routes
5. Zero boarding, daily five-minute TLOS, stops served by any number of routes

Full range of boarding, daily five-minute TLOS, stops served by any number of routes

```
POISSON
; Lhs = round_sa
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,sattlos,
    pfactor,p0xsat,j0xsat,o_stops $

 ************************************************************************
 * NOTE: Deleted    349 observations with missing data. N is now   2388 *
 ************************************************************************


+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:05PM.|
| Dependent variable             ROUND_SA     |
| Weighting variable                 None     |
| Number of observations             2388     |
| Iterations completed                  7     |
| Log likelihood function       -11416.57     |
| Restricted log likelihood     -19760.92     |
| Chi squared                    16688.70     |
| Degrees of freedom                   13     |
| Prob[ChiSqd > value] =         .0000000     |
| Chi- squared = 39142.64050  RsqP=   .8485   |
| G  - squared = 18303.07304  RsqD=   .4769   |
| Overdispersion tests: g=mu(i)  :  3.866     |
| Overdispersion tests: g=mu(i)^2:  2.387     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant    -.6621691704   .91632200E-01   -7.226   .0000
 MEDINC   -.1099409892E-02  .11329304E-02    -.970   .3318     34.194610
 ROAD_JOB  .1047409641E-03  .57383816E-05   18.253   .0000     401.79490
 VEH0HH   -.1402954137E-02  .34587677E-03   -4.056   .0000     23.149322
 AGE18_S     -.7200548035       .13803802   -5.216   .0000     .24090984
 AGEO_S       .5722979317   .64273168E-01    8.904   .0000     .60404811
 FEM_S        1.449997238       .10968472   13.220   .0000     .51047029
 HISP_S       5.863698451       .34034674   17.229   .0000  .32547109E-01
 WHITE_S     -.6546850135   .50546460E-01  -12.952   .0000     .49427571
 SATTLOS   .4241150101E-01  .94596411E-03   44.834   .0000     25.430905
 PFACTOR   .2074246479E-02  .24488340E-01     .085   .9325     .65482777
 P0XSAT    .3699928525E-02  .47088013E-03    7.857   .0000     22.701376
 J0XSAT    .7337526114E-02  .53075541E-03   13.825   .0000     45.396897
 O_STOPS     -.1227057091   .48226184E-02  -25.444   .0000     4.2839196
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

Full range of boarding, frequency, stops served by any number of routes

```
POISSON
; Lhs = round_sa
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,arriv_sa,
    pfactor,p0xsat,j0xsat,o_stops $

 ************************************************************************
 * NOTE: Deleted    369 observations with missing data. N is now   2368 *
 ************************************************************************


+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:05PM.|
| Dependent variable             ROUND_SA     |
| Weighting variable                 None     |
| Number of observations             2368     |
| Iterations completed                  7     |
| Log likelihood function       -12287.44     |
| Restricted log likelihood     -19671.46     |
| Chi squared                    14768.02     |
| Degrees of freedom                   13     |
| Prob[ChiSqd > value] =         .0000000     |
| Chi- squared = 49839.38853  RsqP=   .8060   |
| G  - squared = 20067.17659  RsqD=   .4239   |
| Overdispersion tests: g=mu(i)  :  3.030     |
| Overdispersion tests: g=mu(i)^2:  1.544     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     .9392199131   .84429613E-01   11.124   .0000
 MEDINC   -.9319003688E-02  .12441407E-02   -7.490   .0000     34.205244
 ROAD_JOB  .8911557495E-04  .57872818E-05   15.399   .0000     401.63522
 VEH0HH   -.4924569044E-03  .36496055E-03   -1.349   .1772     23.240138
 AGE18_S     -2.437759557       .14189658  -17.180   .0000     .24113606
 AGEO_S       .4752782273   .66866024E-01    7.108   .0000     .60456679
 FEM_S        1.346284524       .11128180   12.098   .0000     .51111855
 HISP_S       5.398790156       .34749529   15.536   .0000  .32366785E-01
 WHITE_S     -1.085928894   .49824616E-01  -21.795   .0000     .49457168
 ARRIV_SA -.1078983107E-02  .31535363E-03   -3.422   .0006     32.789696
 PFACTOR     -.1419797486   .24305997E-01   -5.841   .0000     .65535683
 P0XSAT    .8995136509E-02  .10957750E-02    8.209   .0000     22.893111
 J0XSAT    .9824227906E-02  .55733740E-03   17.627   .0000     45.780317
 O_STOPS  -.2783628357E-02  .41033468E-02    -.678   .4975     4.2913851
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

No more than 10 boarding, daily five-minute TLOS, stops served by any number of routes

```
POISSON
; Lhs = round_sa<=10
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,sattlos,
    pfactor,p0xsat,j0xsat,o_stops $


 ************************************************************************
 * NOTE: Deleted    346 observations with missing data. N is now   2164 *
 ************************************************************************


+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:06PM.|
| Dependent variable             ROUND_SA     |
| Weighting variable                 None     |
| Number of observations             2164     |
| Iterations completed                  7     |
| Log likelihood function       -4374.493     |
| Restricted log likelihood     -4844.337     |
| Chi squared                    939.6880     |
| Degrees of freedom                   13     |
| Prob[ChiSqd > value] =         .0000000     |
| Chi- squared =  6059.70123  RsqP=   .1199   |
| G  - squared =  5311.60383  RsqD=   .1503   |
| Overdispersion tests: g=mu(i)  : 14.641     |
| Overdispersion tests: g=mu(i)^2: 12.087     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant    -.9155540202       .18028707   -5.078   .0000
 MEDINC   -.3493265936E-02  .18261422E-02   -1.913   .0558     34.765934
 ROAD_JOB -.1303114659E-04  .18399303E-04    -.708   .4788     315.77531
 VEH0HH    .3097327372E-02  .55026621E-03    5.629   .0000     22.175661
 AGE18_S      .5480214056       .25271190    2.169   .0301     .24401630
 AGEO_S       .5766969207       .16226774    3.554   .0004     .60323367
 FEM_S        .4804199337       .24548634    1.957   .0503     .51251278
 HISP_S       5.424995631       .52152663   10.402   .0000  .32507963E-01
 WHITE_S     -.3952804933   .81965172E-01   -4.823   .0000     .50092198
 SATTLOS   .1126776473E-01  .16983440E-02    6.635   .0000     23.873845
 PFACTOR   .4613353838E-01  .38105652E-01    1.211   .2260     .65235441
 P0XSAT    .1690221309E-01  .17413938E-02    9.706   .0000     20.835350
 J0XSAT    .1313461878E-02  .97556931E-03    1.346   .1782     43.388275
 O_STOPS   .5788820009E-03  .87712233E-02     .066   .9474     4.1321627
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

No more than 30 boarding, daily five-minute TLOS, stops served by any number of routes

```
POISSON
; Lhs = round_sa<=30
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,sattlos,
    pfactor,p0xsat,j0xsat,o_stops $

 ************************************************************************
 * NOTE: Deleted    348 observations with missing data. N is now   2345 *
 ************************************************************************


+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:06PM.|
| Dependent variable             ROUND_SA     |
| Weighting variable                 None     |
| Number of observations             2345     |
| Iterations completed                  7     |
| Log likelihood function       -7550.511     |
| Restricted log likelihood     -8816.371     |
| Chi squared                    2531.721     |
| Degrees of freedom                   13     |
| Prob[ChiSqd > value] =         .0000000     |
| Chi- squared = 16092.38799  RsqP=   .1454   |
| G  - squared = 10825.67669  RsqD=   .1895   |
| Overdispersion tests: g=mu(i)  : 10.401     |
| Overdispersion tests: g=mu(i)^2:  8.146     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant    -.3702796128       .12410222   -2.984   .0028
 MEDINC   -.3844626065E-02  .13558342E-02   -2.836   .0046     34.413255
 ROAD_JOB -.3733137761E-05  .11695942E-04    -.319   .7496     341.48353
 VEH0HH    .1126444687E-02  .42836967E-03    2.630   .0085     22.808584
 AGE18_S   .9513428990E-01      .17149744     .555   .5791     .24278530
 AGEO_S       .2420891342   .99905870E-01    2.423   .0154     .60352788
 FEM_S        .6899993775       .16286530    4.237   .0000     .51217162
 HISP_S       5.583207212       .39345197   14.190   .0000  .32686093E-01
 WHITE_S     -.2881723892   .60938511E-01   -4.729   .0000     .49691784
 SATTLOS   .2103403839E-01  .12094672E-02   17.391   .0000     24.883156
 PFACTOR   .1599621635E-01  .28220760E-01     .567   .5708     .65227910
 P0XSAT    .1843846345E-01  .11760937E-02   15.678   .0000     21.880930
 J0XSAT   -.2739831056E-03  .71858645E-03    -.381   .7030     44.589376
 O_STOPS  -.2342510307E-01  .61604127E-02   -3.803   .0001     4.2144989
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

Zero boarding, daily five-minute TLOS, stops served by any number of routes

```
Probit
; Lhs = round_sa=0
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,sattlos,
    pfactor,p0xsat,j0xsat,o_stops $

 ************************************************************************
 * NOTE: Deleted    392 observations with missing data. N is now   2345 *
 ************************************************************************

Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Binomial Probit Model                       |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:07PM.|
| Dependent variable              ZERO_SA     |
| Weighting variable                 None     |
| Number of observations             2345     |
| Iterations completed                  6     |
| Log likelihood function       -1387.291     |
| Restricted log likelihood     -1567.968     |
| Chi squared                    361.3538     |
| Degrees of freedom                   13     |
| Prob[ChiSqd > value] =         .0000000     |
| Hosmer-Lemeshow chi-squared =   3.42258     |
| P-value=  .90511 with deg.fr. =       8     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Index function for probability
 Constant     1.110226493       .25583196    4.340   .0000
 MEDINC    .2119905277E-02  .25746429E-02     .823   .4103     34.413255
 ROAD_JOB -.6734470670E-04  .50582677E-04   -1.331   .1831     341.48353
 VEH0HH   -.4167977824E-02  .12262887E-02   -3.399   .0007     22.808584
 AGE18_S   .2534143068E-01      .44353001     .057   .9544     .24278530
 AGEO_S      -.5434731894       .25389960   -2.141   .0323     .60352788
 FEM_S       -.9775378917       .39227172   -2.492   .0127     .51217162
 HISP_S      -5.176718403       1.0902886   -4.748   .0000  .32686093E-01
 WHITE_S      .5003814461       .13442274    3.722   .0002     .49691784
 SATTLOS  -.1504652796E-01  .31720590E-02   -4.743   .0000     24.883156
 PFACTOR      .1250361115   .63389470E-01    1.973   .0486     .65227910
 P0XSAT   -.2124361375E-01  .33890218E-02   -6.268   .0000     21.880930
 J0XSAT    .2234527128E-03  .16466097E-02     .136   .8921     44.589376
 O_STOPS   .2178241728E-01  .16542999E-01    1.317   .1879     4.2144989
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

WEEKDAY RUNS

Ten test runs are shown for weekday boarding. These vary in the range of boarding, in the measure of service supply, and in the number of routes serving a TLOS stop. These runs are:

1. Full range of boarding, TLOS components, stops served by any number of routes

2. Full range of boarding, TLOS components, two-route stops

3. Full range of boarding, TLOS components, one-route stops

4. No more than 10 boarding, TLOS components, stops served by any number of routes

5. No more than 30 boarding, TLOS components, stops served by any number of routes

6. 10 to 30 boarding, TLOS components, stops served by any number of routes

7. Full range of boarding, frequency, stops served by any number of routes

8. No more than 10 boarding, frequency, stops served by any number of routes

9. No more than 30 boarding, frequency, stops served by any number of routes

10. Zero boarding, TLOS components, stops served by any number of routes

Full range of boarding, TLOS components, stops served by any number of routes

```
POISSON
; Lhs = round_wd
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,
dly_tlos,ndy_tlos,pfactor,p0xwd,j0xwd,o_stops $

 ************************************************************************
 * NOTE: Deleted    169 observations with missing data. N is now   2568 *
 ************************************************************************

+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:08PM.|
| Dependent variable                 ROUND    |
| Weighting variable                  None    |
| Number of observations              2568    |
| Iterations completed                   7    |
| Log likelihood function       -18085.81     |
| Restricted log likelihood     -39063.05     |
| Chi squared                    41954.47     |
| Degrees of freedom                    14    |
| Prob[ChiSqd > value] =         .0000000     |
| Chi- squared = 54359.57480  RsqP=   .9039   |
| G  - squared = 29034.76065  RsqD=   .5910   |
| Overdispersion tests: g=mu(i)  :  4.787     |
| Overdispersion tests: g=mu(i)^2:  2.633     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     .2251716706   .62677928E-01    3.593   .0003
 MEDINC   -.4373061654E-02  .80114751E-03   -5.458   .0000     35.146821
 ROAD_JOB  .1137546650E-03  .36046867E-05   31.557   .0000     390.50293
 VEH0HH    .2889526238E-02  .20037344E-03   14.421   .0000     22.059123
 AGE18_S     -.5881208820   .92688129E-01   -6.345   .0000     .23961809
 AGEO_S       .5001407010   .39273848E-01   12.735   .0000     .60638048
 FEM_S        .5482167563   .70866576E-01    7.736   .0000     .50963484
 HISP_S       5.290646774       .23795601   22.234   .0000  .33213403E-01
 WHITE_S     -.5090066262   .34395205E-01  -14.799   .0000     .50724984
 DLY_TLOS  .4603765448E-01  .65098080E-03   70.720   .0000     20.128505
 NDY_TLOS  .1930522244E-01  .74304580E-03   25.981   .0000     17.370327
 PFACTOR      .1191226925   .16878028E-01    7.058   .0000     .65765773
 P0XWD     .2435706081E-02  .21869358E-03   11.138   .0000     26.312773
 J0XWD     .2775933472E-02  .32534104E-03    8.532   .0000     50.263473
 O_STOPS  -.7984002378E-01  .29713075E-02  -26.870   .0000     4.1573209
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

Full range of boarding, TLOS components, two-route stops

```
POISSON
; Lhs = round_wd
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,
dly_tlos,ndy_tlos,pfactor,p0xwd,j0xwd,o_stops $


 ************************************************************************
 * NOTE: Deleted    152 observations with missing data. N is now   2063 *
 ************************************************************************


+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:08PM.|
| Dependent variable                 ROUND    |
| Weighting variable                  None    |
| Number of observations              2063    |
| Iterations completed                   7    |
| Log likelihood function       -9985.041     |
| Restricted log likelihood     -12091.43     |
| Chi squared                    4212.776     |
| Degrees of freedom                    14    |
| Prob[ChiSqd > value] =         .0000000     |
| Chi- squared = 26299.96635  RsqP=   .2081   |
| G  - squared = 14794.13078  RsqD=   .2216   |
| Overdispersion tests: g=mu(i)  :  6.408     |
| Overdispersion tests: g=mu(i)^2:  5.325     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     .8360259336       .10097086    8.280   .0000
 MEDINC   -.2389968509E-02  .11134262E-02   -2.146   .0318     36.437114
 ROAD_JOB  .1114021290E-03  .20888699E-04    5.333   .0000     217.49454
 VEH0HH    .3587923227E-02  .35765079E-03   10.032   .0000     19.773037
 AGE18_S      .8502759343       .16408762    5.182   .0000     .25175105
 AGEO_S       1.115307863       .10007115   11.145   .0000     .60596118
 FEM_S       -1.943424817       .16433909  -11.826   .0000     .51752069
 HISP_S       4.666665595       .29511090   15.813   .0000  .35123910E-01
 WHITE_S     -.6269424555   .51968178E-01  -12.064   .0000     .52622936
 DLY_TLOS  .2399898510E-01  .11926744E-02   20.122   .0000     16.329132
 NDY_TLOS  .4720025320E-02  .11204101E-02    4.213   .0000     16.264178
 PFACTOR   .5783008663E-01  .22643820E-01    2.554   .0107     .66661780
 P0XWD     .2358457609E-01  .10784501E-02   21.869   .0000     21.099540
 J0XWD    -.8486156375E-03  .53036251E-03   -1.600   .1096     44.994975
 O_STOPS  -.1105376356E-01  .57088192E-02   -1.936   .0528     3.7072225
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

Full range of boarding, TLOS components, one-route stops

```
POISSON
; Lhs = round_wd
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,
dly_tlos,ndy_tlos,pfactor,p0xwd,j0xwd,o_stops $


 ************************************************************************
 * NOTE: Deleted    112 observations with missing data. N is now   1392 *
 ************************************************************************


+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:09PM.|
| Dependent variable                 ROUND    |
| Weighting variable                  None    |
| Number of observations              1392    |
| Iterations completed                   7    |
| Log likelihood function       -5874.381     |
| Restricted log likelihood     -6790.747     |
| Chi squared                    1832.731     |
| Degrees of freedom                    15    |
| Prob[ChiSqd > value] =         .0000000     |
| Chi- squared = 14862.55736  RsqP=   .1891   |
| G  - squared =  8571.38604  RsqD=   .1762   |
| Overdispersion tests: g=mu(i)  :  5.085     |
| Overdispersion tests: g=mu(i)^2:  4.668     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     .2970019222       .13181328    2.253   .0242
 MEDINC    .4312332857E-02  .13523343E-02    3.189   .0014     38.063083
 ROAD_JOB  .1110431137E-03  .40767089E-04    2.724   .0065     191.22639
 VEH0HH    .3397597557E-02  .46959395E-03    7.235   .0000     17.674528
 AGE18_S      1.929703072       .23261082    8.296   .0000     .25374848
 AGEO_S       .6003067522       .15173044    3.956   .0001     .60345288
 FEM_S       -2.003727962       .23549114   -8.509   .0000     .51521841
 HISP_S       7.319007792       .43269376   16.915   .0000  .36576399E-01
 WHITE_S     -.6398817315   .72580898E-01   -8.816   .0000     .54843960
 DLY_TLOS  .3585056334E-01  .18760365E-02   19.110   .0000     12.984914
 NDY_TLOS  .2018577219E-01  .14091678E-02   14.325   .0000     15.484914
 PFACTOR      .1922855978   .31887970E-01    6.030   .0000     .69278985
 P0XWD     .1788623845E-01  .17263996E-02   10.360   .0000     18.349338
 JU0XWD    .2040575833E-02  .96996723E-03    2.104   .0354     21.478256
 JD0XWD   -.4083252900E-02  .94871552E-03   -4.304   .0000     19.414059
 O_STOPS  -.6049161353E-01  .83886678E-02   -7.211   .0000     3.4375000
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

No more than 10 boarding, TLOS components, stops served by any number of routes

```
POISSON
; Lhs = round_wd<=10
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,
dly_tlos,ndy_tlos,pfactor,p0xwd,j0xwd,o_stops $

 ************************************************************************
 * NOTE: Deleted    163 observations with missing data. N is now   2069 *
 ************************************************************************


+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:09PM.|
| Dependent variable                 ROUND    |
| Weighting variable                  None    |
| Number of observations              2069    |
| Iterations completed                   7    |
| Log likelihood function       -4847.327     |
| Restricted log likelihood     -5333.530     |
| Chi squared                    972.4064     |
| Degrees of freedom                    14    |
| Prob[ChiSqd > value] =         .0000000     |
| Chi- squared =  5084.08599  RsqP=   .1249   |
| G  - squared =  5052.52019  RsqD=   .1614   |
| Overdispersion tests: g=mu(i)  : 17.383     |
| Overdispersion tests: g=mu(i)^2: 14.536     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     .1199224759       .12844159     .934   .3505
 MEDINC   -.8140838608E-02  .13837226E-02   -5.883   .0000     36.582783
 ROAD_JOB -.3377671189E-04  .20409719E-04   -1.655   .0979     268.28503
 VEH0HH    .1348717081E-02  .49663888E-03    2.716   .0066     18.936127
 AGE18_S   .9662394682E-01      .19737460     .490   .6245     .24115419
 AGEO_S       .2979790567       .11619691    2.564   .0103     .60436258
 FEM_S        .4563335075       .18339468    2.488   .0128     .51164824
 HISP_S       3.200789487       .49080922    6.521   .0000  .33030472E-01
 WHITE_S     -.2876301688   .65026807E-01   -4.423   .0000     .53341622
 DLY_TLOS  .1709274280E-01  .16194893E-02   10.554   .0000     17.580957
 NDY_TLOS  .8749613595E-02  .14831118E-02    5.899   .0000     17.031899
 PFACTOR   .4819441384E-01  .30799202E-01    1.565   .1176     .65397845
 P0XWD     .6177952764E-02  .13242543E-02    4.665   .0000     22.781912
 J0XWD     .2630738066E-02  .69066982E-03    3.809   .0001     46.851892
 O_STOPS  -.7711001262E-02  .75325814E-02   -1.024   .3060     3.8733688
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

No more than 30 boarding, TLOS components, stops served by any number of routes

```
POISSON
; Lhs = round_wd<=30
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,
dly_tlos,ndy_tlos,pfactor,p0xwd,j0xwd,o_stops $


 ************************************************************************
 * NOTE: Deleted    166 observations with missing data. N is now   2444 *
 ************************************************************************


+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:10PM.|
| Dependent variable                 ROUND    |
| Weighting variable                  None    |
| Number of observations              2444    |
| Iterations completed                   7    |
| Log likelihood function       -9382.694     |
| Restricted log likelihood     -11342.03     |
| Chi squared                    3918.663     |
| Degrees of freedom                    14    |
| Prob[ChiSqd > value] =         .0000000     |
| Chi- squared = 14865.25019  RsqP=   .2294   |
| G  - squared = 12369.96585  RsqD=   .2406   |
| Overdispersion tests: g=mu(i)  : 16.736     |
| Overdispersion tests: g=mu(i)^2: 14.335     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant    -.1974763042       .10064525   -1.962   .0498
 MEDINC   -.3356270059E-02  .97645922E-03   -3.437   .0006     35.662464
 ROAD_JOB  .3036682095E-04  .95129909E-05    3.192   .0014     308.87508
 VEH0HH    .3114581867E-02  .30746072E-03   10.130   .0000     21.106435
 AGE18_S      .7959514823       .13504899    5.894   .0000     .24226324
 AGEO_S       .4559955155   .86466420E-01    5.274   .0000     .60689667
 FEM_S        .6857934719       .13345368    5.139   .0000     .51142392
 HISP_S       5.175137427       .28694466   18.035   .0000  .33614317E-01
 WHITE_S     -.3304728016   .46345312E-01   -7.131   .0000     .51488408
 DLY_TLOS  .2255522211E-01  .10633738E-02   21.211   .0000     19.103110
 NDY_TLOS  .8565441255E-02  .10330740E-02    8.291   .0000     17.285597
 PFACTOR   .3005620694E-01  .20926972E-01    1.436   .1509     .65257208
 P0XWD     .9700333291E-02  .82212141E-03   11.799   .0000     24.601074
 J0XWD     .1868876898E-02  .46634610E-03    4.007   .0001     48.820361
 O_STOPS  -.8139701135E-02  .47343293E-02   -1.719   .0856     4.0278232
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

10 to 30 boarding, TLOS components, stops served by any number of routes

```
POISSON
; Lhs = round_wd
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,
dly_tlos,ndy_tlos,pfactor,p0xwd,j0xwd,o_stops $


 ************************************************************************
 * NOTE: Deleted      4 observations with missing data. N is now    424 *
 ************************************************************************


+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:10PM.|
| Dependent variable                 ROUND    |
| Weighting variable                  None    |
| Number of observations               424    |
| Iterations completed                   5    |
| Log likelihood function       -1340.816     |
| Restricted log likelihood     -1373.593     |
| Chi squared                    65.55550     |
| Degrees of freedom                   14     |
| Prob[ChiSqd > value] =         .0000000     |
| Chi- squared =   754.66269  RsqP=   .0840   |
| G  - squared =   724.64275  RsqD=   .0830   |
| Overdispersion tests: g=mu(i)  :  7.712     |
| Overdispersion tests: g=mu(i)^2:  7.700     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     2.238416627       .15970841   14.016   .0000
 MEDINC    .3012978283E-02  .14960231E-02    2.014   .0440     30.636749
 ROAD_JOB  .3393032571E-04  .11131597E-04    3.048   .0023     531.08770
 VEH0HH    .8250857321E-03  .43257775E-03    1.907   .0565     32.000702
 AGE18_S      .4523893655       .18238965    2.480   .0131     .24493439
 AGEO_S   -.1040708608E-01      .13272816    -.078   .9375     .62109002
 FEM_S     .8217999770E-01      .19076077     .431   .6666     .51119524
 HISP_S       .6159434640       .38516597    1.599   .1098  .36205360E-01
 WHITE_S   .6195263391E-01  .67097985E-01     .923   .3558     .42255998
 DLY_TLOS  .3628221263E-02  .15579710E-02    2.329   .0199     27.259434
 NDY_TLOS  .2925962245E-02  .14405848E-02    2.031   .0422     18.811321
 PFACTOR  -.1036588294E-01  .27500951E-01    -.377   .7062     .65507477
 P0XWD     .5918777744E-03  .11490086E-02     .515   .6065     34.151773
 J0XWD     .1106288689E-02  .61074446E-03    1.811   .0701     59.187832
 O_STOPS   .2955833424E-02  .59726886E-02     .495   .6207     4.8325472
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

Full range of boarding, frequency, stops served by any number of routes

```
POISSON
; Lhs = round_wd
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,arriv_wd,
pfactor,p0xwd,j0xwd,o_stops $


 ************************************************************************
 * NOTE: Deleted    177 observations with missing data. N is now   2560 *
 ************************************************************************


+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:11PM.|
| Dependent variable                 ROUND    |
| Weighting variable                  None    |
| Number of observations              2560    |
| Iterations completed                   7    |
| Log likelihood function       -20435.32     |
| Restricted log likelihood     -39008.62     |
| Chi squared                    37146.59     |
| Degrees of freedom                    13    |
| Prob[ChiSqd > value] =         .0000000     |
| Chi- squared = 69839.19271  RsqP=   .8762   |
| G  - squared = 33745.13989  RsqD=   .5240   |
| Overdispersion tests: g=mu(i)  :  4.037     |
| Overdispersion tests: g=mu(i)^2:  1.904     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     1.809871776   .53294127E-01   33.960   .0000
 MEDINC   -.1063796063E-01  .82867242E-03  -12.837   .0000     35.121648
 ROAD_JOB  .1060150750E-03  .35327559E-05   30.009   .0000     391.22843
 VEH0HH    .2887926104E-02  .21652970E-03   13.337   .0000     22.090425
 AGE18_S     -2.129009674   .92866563E-01  -22.925   .0000     .23964649
 AGEO_S       .4911108642   .40604892E-01   12.095   .0000     .60627317
 FEM_S        .7963327408   .71715093E-01   11.104   .0000     .50962694
 HISP_S       3.255525123       .24098209   13.509   .0000  .33105167E-01
 WHITE_S     -.9364441790   .33663680E-01  -27.818   .0000     .50673375
 ARRIV_WD  .4498389944E-03  .11502794E-03    3.911   .0001     65.199609
 PFACTOR  -.2971709161E-01  .16610098E-01   -1.789   .0736     .65744413
 P0XWD     .2715428640E-03  .53906814E-03     .504   .6145     26.395000
 J0XWD     .9603387257E-02  .32150159E-03   29.870   .0000     50.420546
 O_STOPS   .5933826183E-02  .26202536E-02    2.265   .0235     4.1621094
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

No more than 10 boarding, frequency, stops served by any number of routes

```
POISSON
; Lhs = round_wd<=10
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,
arrive_wd,pfactor,p0xwd,j0xwd,o_stops $

 ************************************************************************
 * NOTE: Deleted    171 observations with missing data. N is now   2061 *
 ************************************************************************


+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:11PM.|
| Dependent variable               ROUND      |
| Weighting variable                None      |
| Number of observations            2061      |
| Iterations completed                 7      |
| Log likelihood function       -4869.096     |
| Restricted log likelihood     -5306.795     |
| Chi squared                    875.3985     |
| Degrees of freedom                   13     |
| Prob[ChiSqd > value] =         .0000000     |
| Chi- squared =  5143.85091  RsqP=   .1090   |
| G  - squared =  5107.42627  RsqD=   .1463   |
| Overdispersion tests: g=mu(i)  : 17.893     |
| Overdispersion tests: g=mu(i)^2: 15.167     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     .4341124472       .12297094    3.530   .0004
 MEDINC   -.8029910707E-02  .13828609E-02   -5.807   .0000     36.557090
 ROAD_JOB -.4244469418E-04  .20194055E-04   -2.102   .0356     268.71177
 VEH0HH    .1579573121E-02  .50110195E-03    3.152   .0016     18.962886
 AGE18_S  -.5248536241E-01      .20053340    -.262   .7935     .24119542
 AGEO_S       .3404889127       .11694637    2.911   .0036     .60422146
 FEM_S        .3972636215       .18408753    2.158   .0309     .51164623
 HISP_S       2.695023440       .49688222    5.424   .0000  .32895321E-01
 WHITE_S     -.4142247758   .63326931E-01   -6.541   .0000     .53287675
 ARRIV_WD  .1803111844E-02  .32705931E-03    5.513   .0000     52.968462
 PFACTOR  -.2795219646E-02  .30339128E-01    -.092   .9266     .65369885
 P0XWD     .9252353389E-02  .13272763E-02    6.971   .0000     22.870342
 J0XWD     .2364735633E-02  .69941904E-03    3.381   .0007     47.033753
 O_STOPS   .2937574649E-01  .61679774E-02    4.763   .0000     3.8782145
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

No more than 30 boarding, frequency, stops served by any number of routes

```
POISSON
; Lhs = round_wd<=30
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,
dly_tlos,ndy_tlos,pfactor,p0xwd,j0xwd,o_stops $


 ************************************************************************
 * NOTE: Deleted    174 observations with missing data. N is now   2436 *
 ************************************************************************


+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:12PM.|
| Dependent variable                 ROUND    |
| Weighting variable                  None    |
| Number of observations              2436    |
| Iterations completed                   7    |
| Log likelihood function        -9545.968    |
| Restricted log likelihood      -11309.54    |
| Chi squared                     3527.152    |
| Degrees of freedom                    13    |
| Prob[ChiSqd > value] =          .0000000    |
| Chi- squared = 15205.44683  RsqP=   .2094   |
| G  - squared = 12707.88136  RsqD=   .2173   |
| Overdispersion tests: g=mu(i)  : 17.229     |
| Overdispersion tests: g=mu(i)^2: 15.285     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     .1933476854   .98151076E-01    1.970   .0488
 MEDINC   -.3485584191E-02  .97894862E-03   -3.561   .0004     35.637703
 ROAD_JOB  .1355274147E-04  .95197017E-05    1.424   .1545     309.36944
 VEH0HH    .3608169615E-02  .30883667E-03   11.683   .0000     21.136202
 AGE18_S      .6069480950       .13673969    4.439   .0000     .24230177
 AGEO_S       .5047264531   .87748543E-01    5.752   .0000     .60678559
 FEM_S        .6595559902       .13475715    4.894   .0000     .51142149
 HISP_S       4.878576558       .29021476   16.810   .0000  .33501889E-01
 WHITE_S     -.4955338119   .45449509E-01  -10.903   .0000     .51436679
 ARRIV_WD  .1746826467E-02  .20735222E-03    8.424   .0000     58.502874
 PFACTOR  -.3273831083E-01  .20569709E-01   -1.592   .1115     .65233090
 P0XWD     .1515310468E-01  .81123656E-03   18.679   .0000     24.681866
 J0XWD     .1470171677E-02  .47356624E-03    3.104   .0019     48.980691
 O_STOPS   .3076063388E-01  .39067355E-02    7.874   .0000     4.0324302
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

Zero boarding, TLOS components, stops served by any number of routes

```
PROBIT
; Lhs = zero
; Rhs = one,medinc,road_job,veh0hh,age18_s,ageO_s,fem_s,hisp_s,white_s,
dly_tlos,ndy_tlos,pfactor,p0xwd,j0xwd,o_stops $


 ************************************************************************
 * NOTE: Deleted    169 observations with missing data. N is now   2568 *
 ************************************************************************

Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Binomial Probit Model                       |
| Maximum Likelihood Estimates                |
| Model estimated: Dec 14, 2004 at 03:30:07PM.|
| Dependent variable                 ZERO     |
| Weighting variable                 None     |
| Number of observations             2568     |
| Iterations completed                  8     |
| Log likelihood function       -1044.817     |
| Restricted log likelihood     -1234.111     |
| Chi squared                    378.5885     |
| Degrees of freedom                   14     |
| Prob[ChiSqd > value] =         .0000000     |
| Hosmer-Lemeshow chi-squared =  10.25737     |
| P-value=  .24742 with deg.fr. =       8     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Index function for probability
 Constant     .2983769157       .27142551    1.099   .2716
 MEDINC    .6003828350E-02  .25147826E-02    2.387   .0170     35.146821
 ROAD_JOB -.8571250320E-04  .95120865E-04    -.901   .3675     390.50293
 VEH0HH   -.4301385641E-02  .16965867E-02   -2.535   .0112     22.059123
 AGE18_S     -.7076282332       .50744620   -1.394   .1632     .23961809
 AGEO_S      -.5044626171       .28595712   -1.764   .0777     .60638048
 FEM_S    -.6060754783E-01      .43610033    -.139   .8895     .50963484
 HISP_S      -4.720220988       1.2315683   -3.833   .0001  .33213403E-01
 WHITE_S      .4788118172       .15314633    3.126   .0018     .50724984
 DLY_TLOS -.1931168404E-01  .46873419E-02   -4.120   .0000     20.128505
 NDY_TLOS -.1253577046E-01  .36709508E-02   -3.415   .0006     17.370327
 PFACTOR   .3118323490E-01  .72202667E-01     .432   .6658     .65765773
 P0XWD    -.1453759024E-01  .39893698E-02   -3.644   .0003     26.312773
 J0XWD    -.4114881987E-02  .16471482E-02   -2.498   .0125     50.263473
 O_STOPS   .2897501042E-01  .20523466E-01    1.412   .1580     4.1573209
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```