Paper Author (s)

Alireza Khani (corresponding), University of Texas, Austin (akhani@utexas.edu)

Tyler James Beduhn, University of Texas, Austin (tyler.beduhn@gmail.com)

Jennifer Duthie, University of Texas, Austin (jduthie@mail.utexas.edu)

Stephen Boyles, University of Texas, Austin (sboyles@mail.utexas.edu)

Ehsan Jafari, University of Texas, Austin (ejafari@utexas.edu)

Paper Title & Number

A Transit Route Choice Model for Application in Dynamic Transit Assignment [ITM # 85]

Abstract

A logit transit route choice model has been estimated using on-board transit survey data. The model is supposed to capture more detailed path attributes and user characteristics such as value of time and trip purpose. The model is suitable for application in a recently-developed schedule-based transit assignment model, compatible with the advanced transportation models.

Statement of Financial Interest

There would be no financial gain for any of the authors from publishing this brief. Modeling results are analyzed from a methodological standpoint, and are not directly related to the characteristics of the selected software tools.

Statement of Innovation

Modeling transit user behavior in a dynamic (schedule-based) network can be a useful tool for planning and decision making. Unlike the current state-of-the-practice transit network models, the proposed transit assignment model provides more detailed information about supply-demand interaction. However, the requirement is that the user behavior is properly studied and used in the assignment. There might be many parameters affecting the transit users' decision making process, leading them to use the paths other than the "shortest path". In this study, we aimed to investigate these parameters, and incorporate them in the dynamic assignment model, so to turn a state-of-the-art model into a practice-ready model. Moreover, the application of the transit data, collected by transit agencies, is investigated for calibration and validation of the assignment model. In general, the model is compatible with the activity-based demand models and dynamic traffic assignment models. Therefore, it can contribute in the next generation of the travel demand models.

A Transit Route Choice Model for Application in Dynamic Transit Assignment

Alireza Khani[1]
Center for Transportation Research
University of Texas at Austin
akhani@utexas.edu

Tyler Beduhn
Department of Civil, Architectural & Environmental Engineering
University of Texas at Austin
tyler.beduhn@utexas.edu

Jennifer Duthie
Center for Transportation Research
University of Texas at Austin
jduthie@mail.utexas.edu

Stephen Boyles
Department of Civil, Architectural & Environmental Engineering
University of Texas at Austin
sboyles@mail.utexas.edu

Ehsan Jafari
Department of Civil, Architectural & Environmental Engineering
University of Texas at Austin
ejafari@utexas.edu

[1] Corresponding Author

1. Introduction

1.1. Background and Motivation

Advanced transportation models, including travel demand models and network supply models, are being more widely used. Nowadays, more planning agencies implement and use models such as activity-based models (ABM) and dynamic traffic assignment (DTA) as powerful tools for planning and decision making purposes. Compatible with these models, an advanced transit assignment model (FAST-TrIPs[2]) has been developed [Khani 2013] and is being implemented in Austin regional area in Texas. The assignment model is capable of modeling schedule-based transit networks and can be integrated with DTA models to capture the interaction between auto and transit networks. Furthermore, it uses a logit route choice model for assigning individual passengers to individual transit vehicles. The route choice model can take into account both path attributes and user characteristics, depending on the behavior of passengers in each application.

In this study, the goal is to calibrate the assignment model for Austin regional area by estimating a route choice model. More specifically, by using existing data including the transit on-board survey and automated passenger count (APC) data, we aim to estimate a transit route choice model to better predict the users' behavior in riding transit vehicles. The route choice model is assumed to follow the logit family structure, and we expect to estimate the parameters of the utility function, including the coefficient of variables such as walking, waiting, and in-vehicle times as well as transfers between routes. Depending on the quality of the data and availability of sufficient information, inclusion of variables such as fare, income, gender, etc. in the model will be investigated.

1.2. Transit Survey Data

The data used in this study were obtained from Capital Metro, the transit agency operating in the Austin region. The data consist of an origin-destination survey conducted during selected weeks from February through May in 2010 by intercept interviews with riders on Capital Metro's fixed route system. The survey determines riders' origin, destination, boarding location to the current route on which they were surveyed, route(s), access and egress travel modes and other demographic information. The stated locations are given as an address or nearest intersection and were geocoded during data preprocessing by Capital Metro. If the respondent indicated transferring to or from the route on which the survey was conducted, only the route number of the connecting bus is known. Of the 32,973 survey records available, 53% (17,587) are weekday records, and approximately 21% of all trips include at least one transfer. Several initial filters are applied to refine the data including eliminating records that have geocoding failures, records that have an identical origin and destination and records that do not include solely walking or transferring as a mode used to access the current bus and the destination. Other access or egress modes such as driving or biking are initially excluded from analysis for simplicity and may be analyzed in future steps. After applying the initial filters, 10,568 potential records are available to infer complete trip attributes.

In the next section, the modeling effort is explained, including the methodology for inferring path attributes from the survey data, generating path choice set using General Transit Feed Specification (GTFS) data, and the model estimation procedure. The final section includes the preliminary results, and a discussion on the application of the model.

[2] Flexible Assignment and Simulation Tool for Transit and Intermodal Passengers

2. Methodology

2.1. Path Inference Model

The first part of the study is to use transit survey data and infer the path attributes for each passenger. For each passenger who participated in the survey, the information about route, boarding location, access and egress mode and origin and destination points are available. However, this information is insufficient for calculating path attributes such as travel time, access or egress time, in-vehicle time, etc. To calculate these attributes, an algorithm was developed that takes survey records, and by comparing them with the transit network (based on GTFS data for the same dates as the survey), finds the most likely path that each passenger has taken. In other words, using the three geographical points of origin, destination, and boarding to the current route, the following information is inferred:

- Boarding and alighting *stops* on each route
- Transfer points between routes, if any
- Walking times including access, egress and transfer(s).

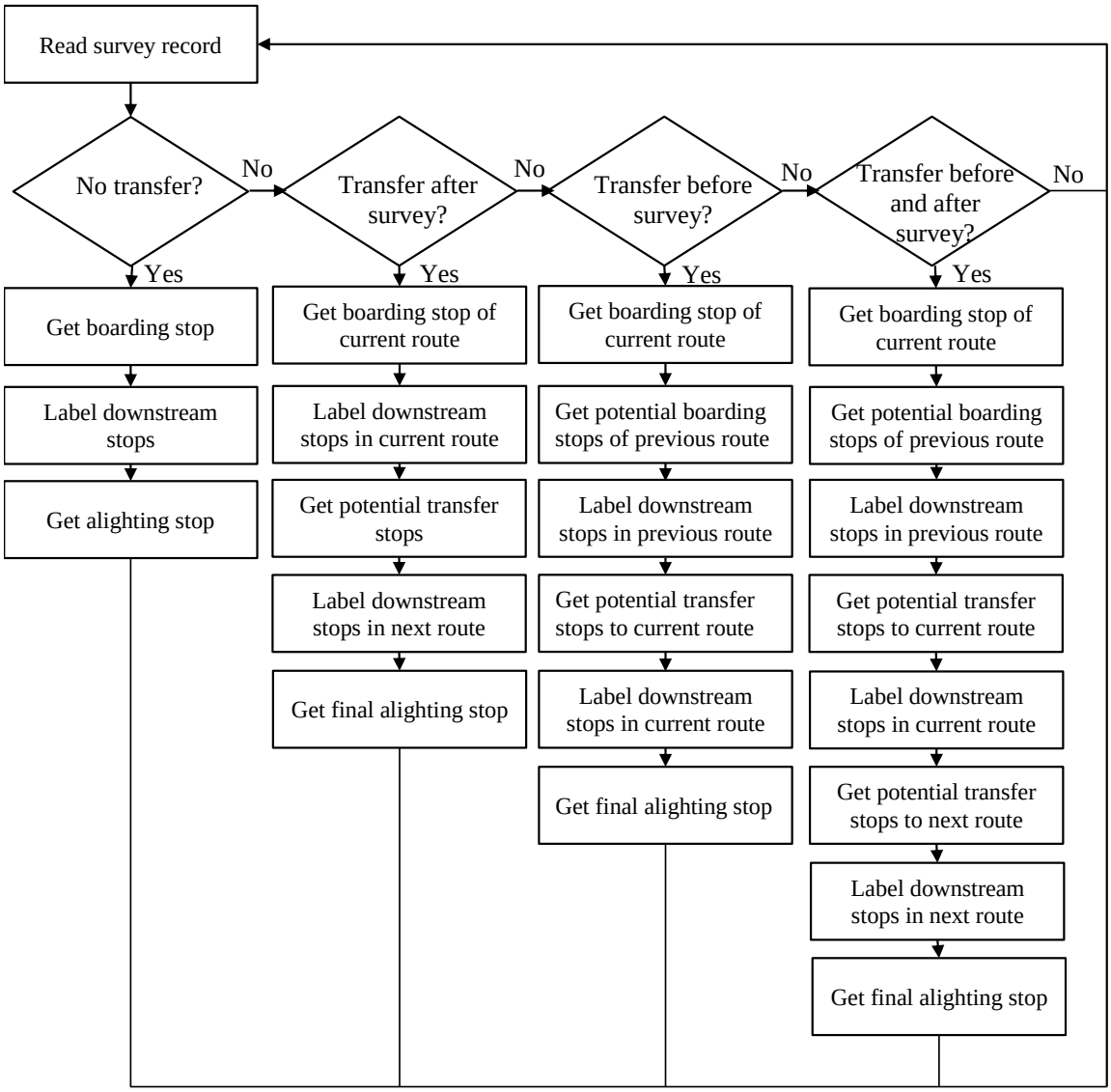

Figure 1. The survey data analysis (path inference) algorithm

The model (shown in Figure 1) uses a constrained shortest path algorithm (using the routes mentioned in the survey) to infer the used path. It also fixes the boarding stop to the route on which the survey is done to the stop closest to the stated boarding location (geocoded intersection or address). A distance label (weighted travel time) and predecessor label is maintained for each stop in a potential path; the path which yields the minimum label at the destination is inferred as the observed path. Depending on the answers to the survey questions, an observed path may include different number of transfers. Thus, four sub-models have been used to analyze the records in the following categories:

- Path with zero transfer
- Path with a transfer before the route on which the survey is done
- Path with a transfer after the route on which the survey is done
- Path with transfers before and after the route on which the survey is done

Several distance thresholds are implemented in the algorithm including a maximum access or egress walking distance of 0.5 mi. and a maximum transfer walking distance of 0.25 mi. These distance filters result in the exclusion of some of observations, but help the inference to be more realistic.

2.2. Choice Set Generation

Generating the set of alternative paths is a challenging problem yet is very critical to the model estimation results. In the route choice problem there is a large number of possible options for a user. Inclusion of all the options in the model estimation process is not only impossible but misleading in estimation of the model parameters. The choice set generation algorithms aim to find the most attractive paths that a user may consider for decision making. Since the final goal in this study is to estimate a model for logit assignment using a hyperpath algorithm, we decided to use a logit-based hyperpath algorithm to generate the set of attractive paths. Hyperpath is a path algorithm that generates a set of elementary paths from an origin to a destination. In other words, in a hyperpath, there may be more than one outbound link (i.e. departing transit vehicle) at each node (i.e. transit stop) with a given travel cost to the destination. The idea is that stops with more attractive transit routes have higher utility, and the utility is calculated by the logsum of the routes' travel cost. Hyperpath has been proposed for application in public transit network modeling [Spiess and Florian 1989 and Nguyen and Pallottino 1988], and has been a useful tool for more realistic traffic and transit assignment. In this study, a currently developed hyperpath model [Khani 2013] is used with modifications to generate the elementary paths for each OD.

Given the origin and destination locations (in Latitude and Longitude) and the time at which the interview was conducted, a hyperpath is generated for each passenger in a time window around the survey time and the unique elementary paths are extracted. The parameters of the utility function are set by a best guess, and the model sensitivity with respect to these parameters are tested later; waiting times are weighted by 2, walking times are weighted by 3 and each transfer is penalized by 5 minutes in addition to the time it takes to make the transfer. The initial waiting time is estimated by its expectation according to another study in the region [Fan and Machemehl 2009]. Finally, the paths with at least 1% probability of being chosen and with at most 2 transfers are kept in the choice set. Note that in the path generation algorithm, multiple choices are involved including the choice of boarding stop, route, transit vehicle (or departure time) and alighting stop in each segment of a trip. Therefore, some generated paths may be very similar and differ in the transit vehicle only. Since the exact boarding time and therefore the exact vehicle taken by a passenger is not known from the survey, the paths that differ in transit vehicle only were combined as a single path. This aggregation resulted in losing some details of a schedule-based network in a small time scale, but since the time window used in the hyperpath model is significantly smaller than

the typical peak and off-peak time periods, the dynamics of the system (i.e. service schedule) is reflected in the model.

2.3. Model Estimation Procedure

After generating the observed paths and the alternatives, the path attributes were calculated and used for model estimation. The variables used in the model, along with their descriptions, are summarized in Table 1. Some user characteristics such as gender, income range, frequent user, type of transit pass, and trip purpose are available from the survey data and can be used in the model. The model estimation was done using BIOGEME version 1.8 [Bierlaire 2003]. In the first step, a multinomial logit model is estimated with the variables whose presence is believed to be significantly important in the utility function, and more complicated model structures may be tested in the future.

Table 1 Variables to be used in the estimation of the route choice utility function

Variable Type	Variable Name	Type	Description
Path Attributes	NTR	Integer	Number of transfers between routes during the path
	IWT	Real (minute)	Initial waiting time (for boarding the first vehicle only)
	IVT	Real (minute)	Sum of the in-vehicle times
	TRT	Real (minute)	Sum of waiting times for making transfers
	TRD	Real (minute)	Sum of walking times for making transfers
	ACT	Real (minute)	Walking time for access to the transit stop from the origin
	EGT	Real (minute)	Walking time for egress to the destination from the transit stop
	LocFare	Real (Dollar)	Sum of fare for the local service rides
	RegFare	Real (Dollar)	Sum of additional fare for the regional service rides
	HW	Real (minute)	Headway of the first route in the path
	Reg	Binary	Indicates whether or not a regional route is used in the path
User Characteristics	Female	Binary (1 if Female)	Gender
	Frequent	Binary (1 if frequent)	Frequent transit user (using transit more than 3 days a week)
	Income	Real (dollar)	Median of income range
	LocPay	Binary (0 if free local ride)	Indicator showing if the passenger has to pay for riding the local services
	RegPay	Binary (0 if free regional ride)	Indicator showing if the passenger has to pay for riding the regional services
	OrigPurpose (i)	Binary for each purpose i	A vector of binary variables indicating the trip's origin purpose
	DestPurpose (i)	Binary for each purpose i	A vector of binary variables indicating the trip's destination purpose

3. Results and Application

3.1. Estimation Results

In the path inference model, after applying the initial filters, 10,568 survey records were processed and a reasonable path was inferred for 6,528 records. The remaining observation records were removed from the data set during the process according to the quality of the data or reasonableness of the inferred path. Possible reasons for why a path was not inferred for all records include error of survey respondents when giving locations or interpreting survey questions, the quality of geocoding and the ability to only capture transfers directly before or after the surveyed route. Below is the summary of the results:

- Trips without transfer: 8,206 records, 5,299 inferred paths (64.6% inference rate)
- Trips with a transfer after the survey: 1,076 records, 578 inferred paths (53.7% inference rate)
- Trips with a transfer before the survey: 990 records, 572 inferred paths (57.8% inference rate)
- Trips with both transfers before and after the survey: 296 records, 79 inferred paths (26.7% inference rate)

The choice set generation model was tested with the default parameters described in section 2.2. An important parameter in the logit model, and therefore in the hyperpath algorithm, is the dispersion (or scale) parameter shown by θ in the following logit model:

$$P(i) = \frac{e^{-\theta u_i}}{\sum_j e^{-\theta u_j}}$$

where u_i is the utility of choice i and $P(i)$ is the probability of choosing choice i. We used the value of 0.5 for θ after doing several sensitivity tests. This setting resulted in average 2.6 paths (with minimum probability of 1%) for each passenger. After removing the observation records for which either the observed path was not generated or the number of paths was less than 2, there remained 2,718 records with average 3.4 paths per record. About 60 percent of these records (i.e. 1655) were randomly selected for model estimation and the remaining were set aside for model validation. Table 2 shows the best estimated multinomial logit model. This model is chosen among several models by doing the common statistical tests on the parameters and the fitness of the model.

Table 2 Parameters of the estimated transit route choice model

Parameter	Value	Standard Error	t-test
β_{IVT}	-0.0733	0.0117	-6.24
β_{IWT}	-0.208	0.0193	-10.76
β_{WALK}	-0.767	0.0981	-7.82
$\beta_{WALK.FREQ}$*	0.230	0.0958	2.40
β_{NTR}	-5.92	0.269	-21.98
β_{TRT}	0.136	0.0483	2.81
$\beta_{PaidFare}$	-0.936	0.413	-2.26
β_{REG}	1.19	0.501	2.37
Number of observations		1,655	
Log-Likelihood with respect to zero		-1892.648	
Final Log-Likelihood		-1052.830	
ρ^2		0.444	
$\bar{\rho}^2$		0.439	

* $\beta_{WALK.FREQ}$: coefficient of the additional utility of walking time by frequent users only

The estimated parameters and their relationship are satisfactory at this stage, given that more complex model specifications are yet to be tested. By normalizing the parameters with respect to the in-vehicle time coefficient, walking time has the weights of 10.46 for general users, and 7.33 for frequent transit users. The initial waiting time has the normalized weight of 2.84 and each transfer has a penalty equal to 80.7 minutes of in-vehicle time (21.1 minutes of waiting time or 7.7 minutes of walking time). It implies that in-vehicle time has a low disutility while transfer has a high disutility. This can be intuitive if we consider factors such as the distribution of observed transfers in the data or the quality and reliability of bus transit. Note that the transfer waiting time has a positive value of 0.136 (1.77 with respect to in-vehicle time). This is because most of the transfers happen within a very short time, implying that people prefer safer transfers (within a range) rather than tight transfers with high risk of missing the next vehicle. This is an interesting finding although it requires more investigation in the future tests. Some other user characteristics and path attributes were used for model estimation and no better model has been estimated as of preparing this paper. Furthermore, a sensitivity test on the parameters of the choice set generation model showed that the estimated model does not change drastically in different settings. The estimated route choice model has been implemented in the schedule-based transit assignment model and the preliminary results were shown to be valid comparing with observed measures through the survey and APC data. However, a complete model validation is necessary, and is being undergone by the authors.

3.2. Application and Future work

The estimated route choice model is an important input to the transit assignment model (FAST-TrIPs in this study). The transit assignment model estimates passenger flow in the transit network, and can be used for ridership analysis, revenue analysis and other planning purposes. It is important to use a behaviorally robust route choice model in the assignment model since the behavior of transit users is generally more complicated than those of other modes, and many parameters in addition to travel time contribute in their decision making. In the estimated model, we tried to capture the effect of parameters such as fare, value of time, and service schedule on transit user behavior. More complex discrete choice models can be tested as well as inclusion of other variables. After running the transit assignment model with the estimated route choice model, the outputs will be compared with the APC data for ridership analyses. The assignment model will also be integrated with a DTA model for development of a dynamic multimodal network model.

4. References

- Bierlaire, M. (2003). BIOGEME: A Free Package for the Estimation of Discrete Choice Models, Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.
- Fan, W. D. and Machemehl, R. B. (2009). Do Transit Users Just Wait for Buses or Wait with Strategies?. Transportation Research Record: Journal of the Transportation Research Board, 2111(1), 169-176.
- Khani, A. (2013). Models and Solution Algorithms for Transit and Intermodal Passenger Assignment (Development of FAST-TrIPs Model), PhD Dissertation. University of Arizona.
- Nguyen, S. Pallottino, S. and Gendreau, M. (1998). Implicit Enumeration of Hyperpaths in a Logit Model for Transit Networks. Transportation Science, 32(1), 54-64.
- Spiess, H. and Florian, M. (1989). Optimal Strategies: A New Assignment Model for Transit Networks. Transportation Research Part B: Methodological, 23(2), 83-102.