INCORPORATING WITHIN-HOUSEHOLD INTERACTIONS INTO A MODE CHOICE MODEL USING A GENETIC ALGORITHM FOR PARAMETER ESTIMATION

Matthew Roorda
Research Associate, Joint Program in Transportation
University of Toronto
35 St. George Street
Toronto M5S 1A4
Canada
Tel: (416) 978-5976
Fax: (416) 978-5054
E-mail: roordam@ecf.utoronto.ca

Eric J. Miller
Bahen Tanenbaum Professor, Department of Civil Engineering
Director, Joint Program in Transportation
University of Toronto
35 St. George Street
Toronto M5S 1A4
Canada
Tel: (416) 978-4076
Fax: (416) 978-5054
E-mail: miller@civ.utoronto.ca

Nicolas Kruchten
Kruchten Engineering Services
377A Dundas Street W.
Toronto M5T 1G6
Canada
Tel: (416) 593-2559
E-mail: nicolas@kruchten.com

Paper presented to the *85th Annual Meeting of the Transportation Research Board.*

National Research Council, Washington D.C., 2006.

Submission Date: November 2, 2005.

Word Count: 5134 words

7 tables & 2 figures = 2250 words

Total = 7384 words

ABSTRACT

This paper describes the procedure for estimating a household model of mode choice. The tour-based mode choice model incorporates inter-personal interactions within the household explicitly in an agent-based random utility modelling framework. Household interactions include vehicle allocation, ride-share to joint activities and drop-off pick-up. Because of the complex nature of the model decision structure, choice probabilities are simulated from direct generation of random utilities rather than through an analytical probability expression. The computational requirements for the simulation are large. Therefore a grid of computers is used in parallel to perform the necessary calculations and a genetic algorithm is used for parameter estimation. This paper presents a brief description of the model, the full model results and a discussion of the computational techniques used in parameter estimation.

INTRODUCTION

This paper describes the procedure for estimating a household model of mode choice. The tour-based mode choice model incorporates inter-personal interactions within the household explicitly in an agent-based random utility modelling framework. Allocation of household vehicles is necessary when the demand for vehicles by members of the household exceeds the number of household vehicles at any time during the day. Vehicles are allocated to maximize total household utility. The passenger mode is modelled as a joint decision between the driver and passenger(s) to ride-share. The model includes explicit evaluation of drop-off and pick-up scenarios, in which the decision to rideshare is made if the utility gain of the passenger exceeds the driver's utility loss (due to making an extra trip or travelling longer distances out-of-the-way).

The model is both tour-based and trip-based. It is trip-based in that the ultimate output of the model is a chosen, feasible travel mode for each trip in the simulation. These trip modes are, however, determined through a tour-based analysis. A key organizing principle in the model is that if a car is to be used on a tour, it must be used for the entire tour, since the car must be returned home at the end. No such constraint, however, exists with respect to other modes such as walk and transit.

Because of the complex nature of the model decision structure, choice probabilities are simulated from direct generation of random utilities rather than through an analytical probability expression. The computational requirements for the simulation are large. Therefore a grid of computers is used in parallel to perform the necessary calculations and a genetic algorithm is used for parameter estimation. The next section of this paper discusses some of the background theory and our motivation for using a genetic algorithm for estimating parameters. This is followed in Section 3 with some definitions, a description of the model method in Section 4 and the data in Section 5. The computational techniques used for parameter estimation are described in Section 6, followed by the full model results and conclusions.

BACKGROUND THEORY

Mode choice modelling has a rich history of well over 30 years in the econometric decision analysis literature. It is one of the classic decisions for which new model structures and theoretical concepts are tested. Common econometric models for modelling tour-based mode choice are multinomial logit (e.g. *1,2*) and nested logit (e.g. *3,4*). These models are based on random utility theory, which assumes that people make rational decisions in order to maximize their level of satisfaction (utility). Utility U(m,t,p) of mode m for trip t on chain c by person p, as shown in Equation 1, is assumed to consist of a systematic component V(m,t,p) which is formulated as a linear function of explanatory variables x and parameters β, and an error term, which is assumed to be randomly distributed.

$$U(m,t,p) = V(m,t,p) + \varepsilon(m,t,p) \qquad t \in T(c,p);\ m \in f(t,p) \qquad (1)$$

where:

V(m,t,p)=	systematic utility component of mode m for trip t for person p
ε(m,t,p)=	random utility component of mode m for trip t for person p
T(c,p) =	set of trips on chain c for person p
f(t,p) =	set of feasible modes for trip t for person p

Under the assumption of random utility maximization, each person p is assumed to choose the mode m that results in the highest U(m,t,p). The distribution of the error term ε(m,t,p) is generally chosen for analytical convenience.
Parameter estimation is usually done by choosing the maximum likelihood set of parameters, (the set of parameters that is most likely to result in the model prediction of observed choices). The log-likelihood function L for a particular set of parameters β is written in Equation 2.

$$L(\beta) = \sum_h \sum_{p\in H(h)} \sum_{c\in C(p)} \sum_{t\in T(c,p)} \log(P(m^*,t,p|\beta)) \quad (2)$$

where:

H(h) = set of persons observed in household h
C(p) = set of home-based tours for person p
β = vector of model parameters (including parameters of the error
P(m*,t,p|β) = simulated probability of person p choosing the observed mode m* for trip t on chain c, given the model parameters β.

Neither the multinomial logit nor the nested logit models, however, are suitable for choice situations where the choice tree structure is non-trivial. Explicit incorporation of vehicle allocation, ridesharing and chain logic in the choice structure can improve the behavioural realism of mode choice models, however, their inclusion results in a level of complexity that does not lend itself well to an analytical solution.

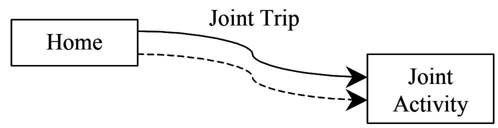
Fortunately, simulation is available as a computationally expensive but flexible tool for maximum likelihood parameter estimation (see *5*). In this technique, the error term is simulated directly, resulting in specific, discrete mode choices being made in the model, rather than an integration of the error terms resulting in a probability for each mode being chosen. The mode choice decision must be replicated in this way many times to achieve a statistically valid representation of the choice process.

The use of a simulation approach allows for the use of virtually any error structure (we have assumed a normal distribution) without adding new layers of complexity to the model estimation process (*6*). It also means that the decision process need not be constrained to mathematical formulations that have a simple closed-form solution. For these two reasons it is chosen for use in this modelling effort, in spite of its computational intensity.

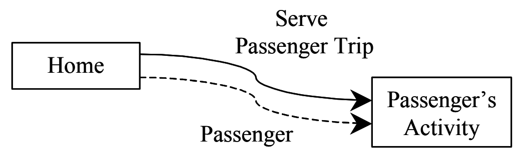
DEFINITIONS

The mode choice model handles different trip and tour types in different ways. In particular there are different scenarios in which household members travel together. To understand the method for handling these scenarios we first present definitions of the different scenarios, in Figure 1.

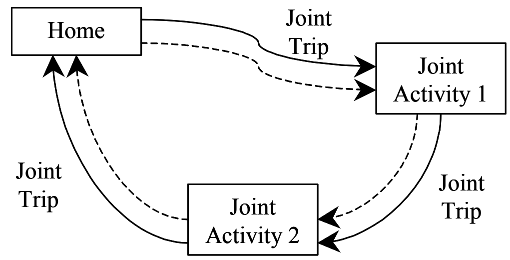
Joint Trip **-** A joint trip is a trip in which more than one household member travel together to or from a joint activity. This can either be a rideshare trip (by car), taking transit together, walking together, etc.



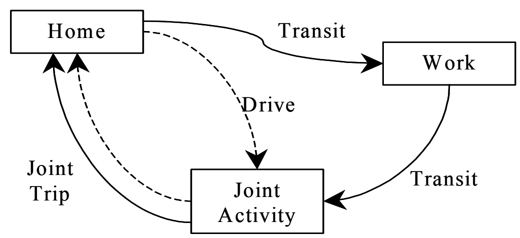
Serve Passenger Trip - A trip made by one member of a household for the purpose of transporting another member to their desired activity. A serve passenger trip may include a passenger (e.g. the trip to drop someone off), or may not include a passenger (e.g. the return trip home after dropping someone off).



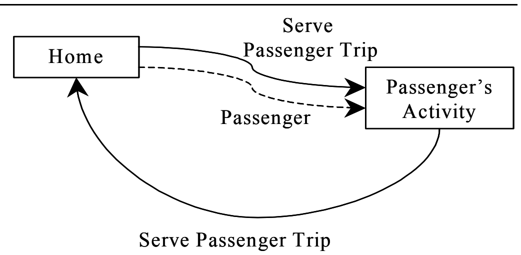
Pure Joint Tour – A joint tour is a tour in which more than one household member travel together to or from at least one joint activity. A pure joint tour occurs when all of the activities on the tours of multiple household members are joint activities. These household members travel together to and from joint activit(ies), and all of these household members make the same trips on the entire tour, at the same times to the same locations. Mode choice for pure joint tours is assumed to be a joint decision, simultaneously determined for all joint activity participants.



Partial Joint Tour - A partial joint tour is possible when some but not all of the activities on the tours of multiple household members are joint activities. A partial joint tour occurs when some but not all of the trips in each of these household members' tours, accessing or egressing from the joint activity, are at the same time, have the same origin and destination, and are by a shared mode.



Pure Serve Passenger Tour - A tour made by one household member solely for the purpose of picking up or dropping off another household member. No activities other than "dropping off" or "picking up" are conducted on a pure serve passenger tour.



En route Serve Passenger Tour - A tour made by one household member that includes at least one serve passenger trip, but also includes other activities before or after the serve passenger trip. For example, a tour in which a parent drops off a child at school on the way to work would be considered an en route serve passenger tour.

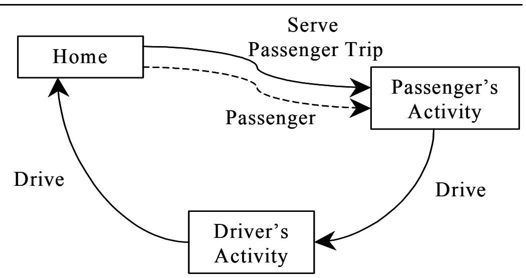


FIGURE 1 Trip and tour definitions

METHOD

The mode choice model method can be summarized as follows:

Step 1 Individual tour mode choice

- Mode choice for individual trip-maker is determined based on a random utility maximization framework that incorporates a tour-based decision tree structure.
- Trip-level and tour-level rules for mode availability are enforced.
- For pure joint tours (tours that involve more than one person for all activities on the tour) mode choice is determined simultaneously by all tour participants.

Step 2 Vehicle allocation

- Allocation decisions are made at the household level to maximize household utility.

Step 3 Serve passenger matching procedure

- Compatible individual tours are considered for *en route* serve passenger tours and partial joint tours. Shared rides are chosen for compatible individual tours if it improves total household travel utility.

Step 4 Pure serve passenger tours

- If *en route* serve passenger opportunities are not available, then pure serve passenger tours are considered if a driver is available at home. Such arrangements are chosen if it improves total household utility.

A conceptual description of the mode choice model presented in this paper has been described in (*6*). However, the prototype implementation of this model was incomplete. It did not include the implementation of joint tours, within household ridesharing, drop-offs and pick-ups. The mode choice model presented in this paper includes the following modes: auto drive, auto passenger (serve passenger tours), rideshare (for pure joint tours), transit all-way, and walk. Bicycle, taxi, drive access transit, commuter rail and school bus are excluded because they represent a very small proportion of total reported trips in the Greater Toronto Area. Carpool (inter-household) trips are excluded from the model and reserved for future study because they require understanding of inter-household interactions. The remainder of this section gives a brief summary of the methods for the steps described above.

Individual Trip-Maker Tour Mode Choice

The objective of the mode choice model is to select a mode of transportation for each trip made by each member of the household such that trip and chain level rules are satisfied, vehicle allocation constraints are not violated and the total utility of the household is maximized.

Consider first the case where an individual chooses the modes of transportation for trips within a tour. Figure 2 shows the basic trip chain-level decision that is made, within which a number of sub-decisions are made for individual trips within the trip chain. Clearly, the choice of whether to take an automobile along on the chain has very strong implications for all trips on that chain and for the decision making of other household members. First, it is necessary that the vehicle be returned back home at the end of the chain. Second, the vehicle must be used for every trip on the chain unless the driver plans to return to the same location to "pick up the vehicle" before returning home. Finally, if the vehicle is being used by one household member, it is unavailable for other members to use as long as it is out of the driveway.

Conceptually, bicycles could be treated in exactly the same way as automobiles, as shown in Figure 2. However, the bicycle mode is not included in the model specification presented in this paper, because the mode comprises only 0.8% of total reported trips in the base data.

For the "non vehicle" chain alternative, the mode choice decision for each individual trip is assumed to be independent of the decisions for the other trips within that chain. The maximum utility non-personal vehicle mode is chosen for each trip in the chain.

Sub-chains add an additional level of complexity to the choice structure. For example, an individual can drive to work, walk to a restaurant for lunch, walk back to work, and return home by car at the end of the workday. Generally, if a sub-chain exists, then non-vehicle modes are available for trips on that sub-chain even if the main part of the chain is made using a car or a bicycle.

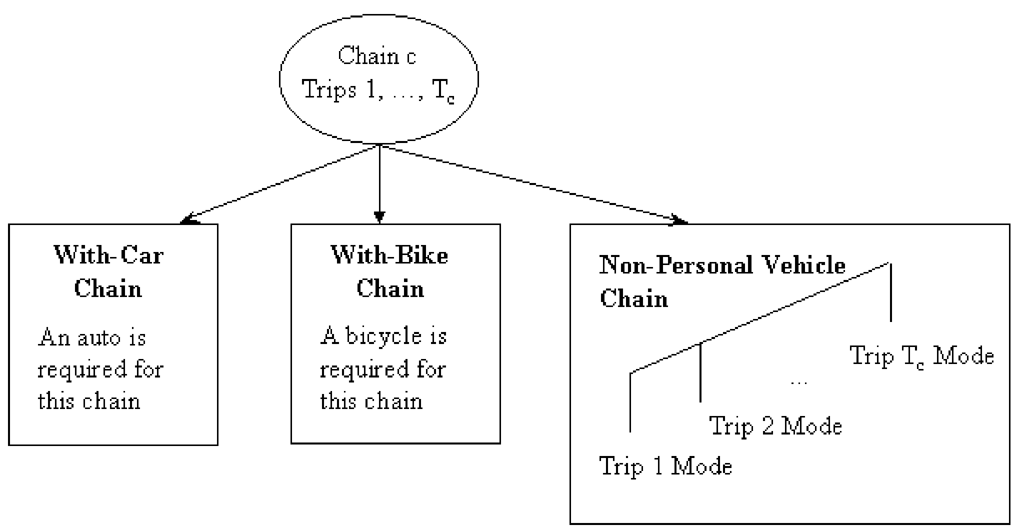


FIGURE 2 Mode choice tour-level decision tree

There are a number of sub-choices within the with-car chain alternative. A car may be used to access a commuter rail station or a park-and-ride subway station. The car may be used to get to work but may be left in the work parking lot while the owner walks or takes a taxi to go for lunch (or in other words, to participate in a work-based sub-chain), only to return to work and drive back home again. These complex chain types are handled in the model using the concept of a "chain mode set", defined as a feasible set of modes for the trips in the chain that satisfy the constraint that an automobile must be returned home at the end of the chain without being left stranded at any point in the chain.

As discussed in (*6*), a random utility approach is adopted in this model to determine the choice among these options. The utility of person p choosing mode m for trip t on trip chain c, U(m,t,p), is formulated in the usual way as shown in Equation 1.

Further, we assume that the utility for a specific combination of modes for the entire trip chain c, U(M,p) is simply the sum of the individual trip utilities shown in Equation 3:

$$U(M,p) = \sum_{t\in T(c,p)} V(m(t),t,p) + \sum_{t\in T(c,p)} \varepsilon(m(t),t,p) \qquad M\in F(c,p) \qquad (3)$$

where:

M = one set of specific feasible modes for the trips on chain c for person p (the *chain mode set*)

F(c,p) = set of chain mode sets for chain c for person p; this set is defined by both *a priori* trip constraints (e.g., trip distance too long to walk) and chain-based "contextual" constraints (e.g., can't use auto-drive on return trip if it was not used on the outbound trip)

Equation 3 is a key assumption in the model design, although it is not clear that attractive, practical alternatives to this assumption exist. It is essential to provide a consistent comparison between chain-based and trip-based modes, as well as to deal with ridesharing and joint-travel mode choices. Also note that this linear additive assumption is implicit in conventional trip-based models.

The standard random utility assumption is made that the chain mode set chosen is M* for which:

$$U(M^*,p) \geq U(M,p) \qquad \forall\, M,M^* \in F(c,p);\ M^* \neq M \qquad (4)$$

That is, the drive or (optimal) non-vehicle chain mode set will be chosen for the given trip chain, depending on which provides the maximum utility to the trip-maker.

We use a microsimulation approach to evaluate Equation 4 directly. For each trip, the ε(m,t,p) term is generated randomly, assuming a normal error distribution. Given the randomly generated ε's, and the systematic utility V(m,t,p) the highest utility chain mode set can be identified and chosen. To determine the maximum likelihood parameters, it is necessary to compute P(M*,p) by replicating the process many times and determining the frequency with which the observed chain mode set is chosen. This is discussed in more detail in Section 6.

Pure Joint Tours

If multiple household members engage in tours that consist only of joint activities, then the tour is labelled a pure joint tour and mode choice is considered to be a joint decision. For pure joint tours, it is assumed that the total utility for all persons involved on the joint tour is equal to the sum of the utilities of the joint tour mode set (M_j) for each person:

$$U(M_j,J)= \sum_{p\in J(c)} U(M_j,p) \qquad M_j\in F_j(c,J(c)) \qquad (5)$$

where:

M_j = one set of specific feasible modes for the trips on pure joint tour (chain) c for person p (a *joint tour mode set)*

$F_j(c,J(c))$ = set of joint tour mode sets for chain c for persons J(c); this set is defined by both *a priori* trip constraints (e.g., trip distance too long to walk), chain-based "contextual" constraints (e.g., can't use auto-drive on return trip if it was not used on the outbound trip) and joint tour constraints (all persons must travel using the same mode of transportation)

J(c) = set of persons participating on the joint tour c

Vehicle Allocation

Initially, the mode choices for individual tours and for pure joint tours are made without regard for the availability of household vehicles at particular times of day. However, in many instances the tours in a household overlap in time. If the number of vehicles available to the household is less than the number of overlapping with-car tours, then a decision must be made as to which household member uses the vehicle. In such cases, all possible vehicle allocations are evaluated and the allocation that results in the highest overall household utility is chosen. Those household members that are not allocated a vehicle as a result of this evaluation are assumed to choose the highest utility non-vehicle chain mode set.

En-Route Serve-Passenger Tours

Opportunities exist within the household for ridesharing, even when activities are not done together. One common example is that of parents dropping-off or picking-up children at school *en-route* to work or other activities. In such situations, the person that serves the passenger (i.e. gives them a ride) experiences an increase in travel time and inconvenience, in order to chauffeur another household member, who as a result experiences an improvement in travel utility. Our model approach assumes that household members first consider their tour mode choices individually. Serve passenger opportunities are then evaluated in terms of total household utility, that is, if the increase in utility experienced by the passenger exceeds the decrease in utility experienced by the driver, then the serve passenger arrangement is chosen. Otherwise, the driver drives alone, and the would-be passenger chooses the best alternative available mode. Formally, the travel utility gain for the serve passenger alternative $\Delta U_{s,p}$ for person p involved on the serve passenger trip can be written as follows:

$$\Delta U_{s,p} = \Sigma_{t_s \in T_s(c,p)} U(m_s,t_s,p) - \Sigma_{t \in T(c,p)} U(m,t,p) \tag{6}$$

where:

t = the trip made without serve passenger
t_s = the trip made with serve passenger
m = the mode for trip t without serve passenger
m_s = the mode for trip t_s with serve passenger
$T(c,p)$ = the set of trips t on chain c for person p, without serve passenger
$T_s(c,p)$ = the set of trips t_s on chain c for person p, with serve passenger

The total household travel utility gain for the serve passenger alternative ΔU_s is:

$$\Delta U_s = \Sigma_{p \in J_s(c)} \Delta U_{s,p} \tag{7}$$

where:

$J_s(c)$ = the set of persons involved in the serve passenger alternative for chain c.

The serve passenger alternative is chosen if $\Delta U_s > 0$.

We note that, in addition to the change in travel utility, there may also be changes to activity utility (the utility that one derives from participating in an activity at some time, at some location, for some duration) because activities can change when those wanting to serve a passenger need to modify their schedules. In this model of mode choice, as with other mode choice models in the literature, this utility is ignored.

Partial Joint Tours

A tour is a candidate to be a partial joint tour if at least one but not all of the out-of-home activities on the tours of multiple persons in the household are joint activities. The procedure for assessing partial joint tours is a special case of the more general procedure described for serve passenger tours. By definition, tours that are candidates to be partial joint tours include trips that have common origin, destination and timing for multiple people. For these trips, therefore, it is not necessary for one person to make adjustments to their tour to include a "drop-off / pick-up" activity. Hence, the rideshare feasibility rules need not be applied in this case; the rideshare mode will always be feasible provided there is a vehicle available. Equations 6 and 7 can therefore be applied directly to determine whether a partial joint tour is formed.

DATA

The mode choice model is based on data from the 1996 Transportation Tomorrow Survey (TTS), a conventional trip diary survey carried out on approximately 5% of the population of the Greater Toronto Area. This high quality data set is described in detail in (*7,8*). In order to support a household chain-based model, a significant effort was made to clean the data, to identify trip chains, to identify and classify joint trips and serve passenger tours, and to attach level of service information not included in the TTS database. Only major modes (including drive, transit all-way, walk, passenger, and rideshare) were retained. Notably, inter-household carpooling is not included within this model.

For model estimation purposes, a sample of households was drawn randomly from the processed TTS dataset. Only trips that conformed to all choice set rules were used for model estimation. If a trip violated the choice set rules, the entire household was removed from the estimation dataset. Table 1 shows a summary of the total sample and the estimation subsample.

PARAMETER ESTIMATION

Calculating the Log-Likelihood

Model parameter values were estimated by maximizing the log-likelihood function shown in Equation 2. Because of the non-standard trip chain "nesting" structure shown in Figure 2, and the complexities of vehicle allocation, ridesharing and serving passengers, no analytical expression could be found for the choice probability P. Thus, P is simulated through a Monte Carlo process in which N sets of random utilities U are drawn for each trip for each person for each chain for a given β, Equation 4 is evaluated for each draw, and the frequency with which m* is predicted to be chosen is accumulated. In order to account for the possibility that m* is never chosen within the N draws, P is defined as shown in Equation 9 (*9*):

$$P(m^*,t,p|\beta) \quad = \quad [F(m^*,t,p|\beta) + 1] / [N + n_t] \qquad (9)$$

where:

$F(m^*,t,p|\beta)$= the number of times m* was selected for trip t out of the N draws, and n_t is the number of feasible modes for trip t.

TABLE 1 1996 TTS Total Sample & Estimation Sub-Sample Summary

	TTS Total Households (Raw)		TTS Processed Households with major modes only		Initial Estimation Set		Final Estimation Set[a]	
Households	88898		45565		4465		4049	
Persons	243286		117404		11446		7154[b]	
Trip Chains	N/A		100706		9073		8603	
Trips	500313		229178		20442		19335	
Drive[c]	311502[d]	62.3%	139819	61.0%	12199	59.7%	11702	60.5%
Transit All-way	59760	11.9%	35846	15.6%	3297	16.1%	3118	16.1%
Walk	29250	5.8%	16128	7.0%	1395	6.8%	1299	6.7%
Passenger (drop-off / pickup)	78768[e]	15.7%	8012	3.5%	650	3.2%	399	2.1%
Rideshare (to joint activities)	N/A		29373	12.8%	2901	14.2%	2817	14.6%
Drive Access Subway	863	0.2%	0	0.0%	0	0.0%	0	0.0%
Drive Egress Subway	798	0.2%	0	0.0%	0	0.0%	0	0.0%
Drive Access Commuter Rail	1241	0.2%	0	0.0%	0	0.0%	0	0.0%
Drive Egress Commuter Rail	1161	0.2%	0	0.0%	0	0.0%	0	0.0%
Non-drive Commuter Rail[f]	2216	0.4%	0	0.0%	0	0.0%	0	0.0%
Taxi	2386	0.5%	0	0.0%	0	0.0%	0	0.0%
School bus	7684	1.5%	0	0.0%	0	0.0%	0	0.0%
Bicycle	3891	0.8%	0	0.0%	0	0.0%	0	0.0%
Other/Unknown	793	0.2%	0	0.0%	0	0.0%	0	0.0%

(a) Trips by modelled modes complying with all choice set rules.
(b) Only includes persons making a trip. Other columns include all persons in the household.
(c) Drive includes motorcycle trips.
(d) In the raw TTS data, rideshare to joint activities is not identified as a separate mode. Rideshare drivers are included in this number.
(e) In the raw TTS data, rideshare to joint activities is not identified as a separate mode. Rideshare passengers are included in this number
(f) "Non-drive Commuter Rail" indicates transit, walk, auto passenger and taxi commuter rail station access modes.

Genetic Algorithm for Maximum Likelihood Parameter Estimation

The search for a parameter set that resulted in the maximization of the log-likelihood function was done using a genetic algorithm (GA). Simply put, genetic algorithms are a method of searching multi-dimensional space according to some criteria (*10*). The method is based on an analogy to the evolutionary process in nature, in which populations of organisms adapt and change over time through processes of selection, reproduction, and mutation. Those organisms with the genetic makeup that is most well adapted to the environment are most likely to survive and reproduce, resulting in a population that is increasingly "fit".

As applied to the problem of mode choice parameter estimation, the genetic algorithm analogy is developed as follows. Each vector of parameters $\underset{\sim}{\beta}$ is considered a *chromosome*, and each individual parameter within that vector is a *gene* within that chromosome. The *fitness* of the chromosome is the value of the log-likelihood function in Equation 2, thus, the chromosome with the highest fitness is the maximum likelihood parameter set.

The GA begins by initializing a *population* of chromosomes, that is, a group of feasible parameter sets (β). This initial population forms the first generation of the evolutionary process. The fitness ($L(\beta)$) of each of those chromosomes is then evaluated. Based on the fitness of each of the chromosomes in the population, a process of *selection* takes place in which chromosomes with higher fitness *survive* and unfit chromosomes are discarded. *Reproduction* involves the selection and the mixing of the genes of two *parent* chromosomes (parameter sets that have survived the selection process) to result in *child* chromosomes. The process of reproduction involves both *recombination*, the mixing of genetic information from the two parents, and *mutation*, the introduction of slight modifications to individual genes in the chromosome. The next generation is then built using an *assembly* step, in which a subset of the parent and child populations are chosen through the process of selection. This evolutionary process repeats itself over many generations, and the overall fitness of the population improves in each generation.

The design of a GA for a particular application involves the choice of methods for each of the processes as described above. Significant testing was undertaken to find a combination of methods that was feasible given our computing resources, converged reasonably quickly to a solution, and was able to find the maximum likelihood parameter set with a good degree of consistency. Table 2 shows the GA methods chosen based on these tests. While this GA configuration is not necessarily optimal, it was sufficient to solve the maximum likelihood parameters for our mode choice application with reasonable efficiency. The choice of an appropriate population size was a critical decision because it directly influenced the computing resources required and the speed of convergence. This setting (as well as the other settings) were tested by applying the genetic algorithm multiple times with different random number streams. To obtain stable estimation results we found that it was necessary to use a population size of approximately twice the number of parameters to be estimated. However, further increasing the population size achieved limited benefits in the consistency of the solution. Hence, for most model runs, a population size of 50 was used, however, in the final set of model runs the population size was increased to 70.

TABLE 2 Genetic Algorithm Methods and Settings

Genetic Algorithm Element	Chosen Method or Setting	Description
Population size	70	
Initial population	Random selection	Genes for each chromosome in the initial population are randomly generated within the defined search space.
Selection (for assembly)	Best Selector	Selects chromosomes with the highest fitness.
Selection (for recombination)	Rank-based Selector	Chromosomes are ranked in order of fitness. Chromosome a is selected with probability defined by: P[a] = (popsize-rank(a))/(popsize).
Assembly	Crowding Assembler	Pools the parent and child chromosomes and selects those with the highest fitness (Best Selector)
Recombination	Real Multi-Crossover	Parents are chosen using the Rank-based selector. Each of the genes of the child is set to the value of that gene in one or the other of the parents, with equal probability.
Mutation	Percentage Gene Mutator	All parameters were mutated within a range of +/- 7%
Stopping Criterion	No. of generations with no increase in max likelihood	15

The GALAPAGOS software (*11,12*) was used to apply the genetic algorithm to the problem of mode choice parameter estimation. GALAPAGOS is built upon the LIGHTGRID grid-computing engine, which allows calculations to be run in a distributed computing environment. Given that an available grid of computers is linked to a common network, LIGHTGRID manages the process of setting up client computers and dispatching computing tasks to those clients. By parallelizing the computing using GALAPAGOS and LIGHTGRID, the speed with which a model could be estimated was dramatically improved. As noted by Kruchten (*11*), the optimal usage of the client computers could be made when:

Population size / number of client computers = integer value (10)

The minimum model run time was therefore obtained by allowing the number of client computers to match the population size. This was feasible for this project, because a large networked computing grid of moderately powerful 1.4 GHz single-processor desktop machines was available for use during off-hours. Time to convergence for the model runs ranged from 1.5 to 2.5 hours using the distributed computing environment.

MODEL RESULTS

Table 3 presents parameter estimates, likelihood ratio tests for these parameters, and goodness-of-fit statistics for the model. Table 4 shows statistics for the explanatory variables. The following results are notable from Table 3:

- All parameters have expected signs.
- Given the parameter estimation procedure used, asymptotic t-statistics cannot be readily computed. Instead, likelihood ratio tests were performed for each parameter by deleting the parameter and re-estimating the model. Using that test, all parameters are strongly significant. One of the least significant variables was that of *travelcost*. We note that the coefficient for this parameter showed some variability between model runs, yet it was retained as a key policy variable.
- All parameters are of plausible magnitude. However, the *travelcost* parameter, which, combined with the parameter values for *atime* and *tivtt*, implies values of time of $69/hr for auto users and $27/hr for transit users. These values of time are somewhat higher than expected.
- The mode choice model fits the data very well and produces a fairly high overall goodness of fit (an adjusted ρ^2 of 0.710).
- An important concern in simulated log-likelihood calculations is the possibility that an observed mode for a given observation is never chosen within the Monte Carlo simulation. In this analysis, 100 random draws were generated per trip. As shown in Table 3, 166 of the 19,335 trips (0.86%) did not have the observed chosen mode selected at least once during the simulation. While ideally this number should be driven to zero as the estimation proceeds, such a small number of "never chosen" trips is not likely to be having a large impact on the model estimation results.
- Of the 166 trips never correctly predicted, 72 were auto passenger trips, comprising 18% of total auto passenger trips. For other modes, less than 2% of total trips were never correctly chosen. A manual review of the base data for these trips uncovered no obvious reason to explain why they were not chosen. However, this result indicates that passenger mode is clearly the

most difficult mode of transport to predict correctly. This is not surprising, because the passenger mode (for drop-offs and pickups), involves a negotiation among household members for which we only have a limited understanding. Furthermore, it impacts not only the trip attributes of the driver and passenger but also the activity schedule of the driver (a drop-off activity is added).

TABLE 3 Model Estimation Results

Parameter	Description	Coeff-icient	Lik. Ratio
c-tr_n_dr	Mode specific constant for transit all-way	-0.166	18.46
c-walk	Mode specific constant for walk	-0.304	28.96
c-ridesh	Mode specific constant for rideshare (for joint trips)	0.835	72.40
c-pass	Mode specific constant for auto passenger	-2.385	527.0
atime	Auto in-vehicle travel time (min)	-0.075	167.2
tivtt	Transit in-vehicle travel time (min)	-0.029	94.7
twalk	Walk travel time including walk access to/from transit (min)	-0.064	1263.5
twait	Transit wait time (min)	-0.145	267.8
travelcost	Travel cost ($1996 Canadian)	-0.065	28.7
pkcost	Parking cost ($1996 Canadian)	-0.302	314.2
dpurp_shop_d	=1 if trip purpose = shopping (drive mode); = 0 otherwise	0.993	174.0
dpurp_sch_d	=1 if trip purpose = school (drive mode); = 0 otherwise	-1.181	302.1
dpurp_oth_d	=1 if trip purpose = other (drive mode); = 0 otherwise	0.593	116.7
dest_pd1_w	=1 for walk trips destined for downtown Toronto; = 0 otherwise	0.897	114.3
intrazonal_t	=1 for an intrazonal trip for transit all-way mode; = 0 otherwise	-2.962	299.9
adjzone_t	=1 for an adjacent zone for transit all-way mode; = 0 otherwise	-1.016	142.2
age11_15_p	=1 if age 11-15 (passenger mode); =0 otherwise	0.954	61.3
Etrip_par	Scaled variance for the trip specific error term	1	
Num Observations		19335	
Num Parameters		17	
Log Likelihood L(β)		-5035.87	
Log Likelihood No Parameters L(0)		-17434.8	
-2[L(0)-L(β)]		24797.8	
rho^2		0.7112	
Adjusted rho^2		0.7102	
Number of Observations in which observed mode never chosen		166	

TABLE 4 Mean, Standard Deviation of Explanatory Variables in Final Estimation Dataset

Variable	Average	Std.Dev.	Variable	Average	Std.Dev.
atime	12.3	11.3	dpurp_shop_d	0.174	0.379
tivtt	25.0	21.7	dpurp_sch_d	0.088	0.284
twalk	21.9	21.4	dpurp_oth_d	0.243	0.429
twait	7.4	5.0	dest_pd1_w	0.077	0.267
travelcost	1.6	1.5	intrazonal_t	0.075	0.263
pkcost	0.76	1.97	adjzone_t	0.033	0.180

Tables 5, 6 and 7 present prediction-success tables for the model. Again, the good fit of the model is indicated in these tables, with over 88% of observed modes being chosen on average. In addition, each mode except for the passenger mode is well predicted with prediction success rates in the order of 95%, 74% and 67%, and 99% for the auto-drive, transit, walk, and rideshare modes, respectively. For these modes, relatively little "confusion" exists within the model, with off-diagonal elements being generally small and "well balanced" (approximately as many transit trips are incorrectly assigned to walk as walk trips are assigned to transit, and so on). However, the prediction success rate of passenger trips is a relatively low value of 21%. Attempts were made to improve this result by including additional model parameters (for example, destination purpose, and additional dummy variables for different age categories, and sex). All of these additional variables were found to be insignificant.

TABLE 5 Prediction Success Table for the Estimated Model (Trips)

	Predicted Mode					
Observed Mode	Drive	Transit	Walk	Rideshare	Passenger	Total
Drive	11054	448	139	0	59	11699
Transit	487	2313	216	28	73	3117
Walk	122	252	863	2	60	1299
Rideshare	0	0	3	2814	14	2831
Passenger	109	125	70	0	81	385
Total	11771	3139	1291	2844	287	19331

TABLE 6 Prediction Success Table for the Estimated Model (% of Total Trips)

	Predicted Mode					
Observed Mode	Drive	Transit	Walk	Rideshare	Passenger	Total
Drive	57.2%	2.3%	0.7%	0.0%	0.3%	60.5%
Transit	2.5%	12.0%	1.1%	0.1%	0.4%	16.1%
Walk	0.6%	1.3%	4.5%	0.0%	0.3%	6.7%
Rideshare	0.0%	0.0%	0.0%	14.6%	0.1%	14.6%
Passenger	0.6%	0.6%	0.4%	0.0%	0.4%	2.0%
Total	60.9%	16.2%	6.7%	14.7%	1.5%	100.0%

TABLE 7 Prediction Success Table for the Estimated Model (% of Observed Mode)

	Predicted Mode					
Observed Mode	Drive	Transit	Walk	Rideshare	Passenger	Total
Drive	94.5%	3.8%	1.2%	0.0%	0.5%	100.0%
Transit	15.6%	74.2%	6.9%	0.9%	2.3%	100.0%
Walk	9.4%	19.4%	66.5%	0.2%	4.6%	100.0%
Rideshare	0.0%	0.0%	0.1%	99.4%	0.5%	100.0%
Passenger	28.3%	32.5%	18.1%	0.0%	21.0%	100.0%
Total	60.9%	16.2%	6.7%	14.7%	1.5%	100.0%

It is also noted in Table 6 that the aggregate predicted mode shares for the auto-drive, transit, walk and rideshare modes very closely match the observed mode shares. However, the total predicted mode share for the passenger mode was found to be 75% of the observed mode share. Unlike a conventional logit model estimation procedure, for example, in which predicted and observed mode shares are forced to match through the selection of the alternative-specific parameter values, no such constraint is imposed within this model's estimation. Thus, the ability to reproduce the observed shares for all but one of the modes is a reasonably strong test of the model's overall performance.

CONCLUSIONS

The household tour-based mode choice model developed for the Greater Toronto Area is a significant improvement on the existing models used in practice in that it incorporates household level interactions explicitly, including vehicle allocation within the household, joint travel decisions, and negotiations over ridesharing in the household. All of the decisions are modelled within a clear theoretical framework of household random utility maximization. It is also demonstrated that the maximum likelihood parameters of this model can be estimated using a Monte Carlo simulation technique for computation of the log likelihood, and that a genetic algorithm in a distributed computing environment can be successfully used for efficient search of parameter space with acceptable model run times.

REFERENCES

1. Bradley, M., M. L. Outwater, N. Jonnalagadda, and E.R. Ruiter. Estimation of Activity-Based Microsimulation Model for San Francisco. In *Proceedings of the 80th Annual Meeting of the Transportation Research Board*, CD-ROM. Transportation Research Board, National Research Council, Washington D.C., 2001.

2. Jonnalagadda, N., J. Freedman, W.A. Davidson, and J.D. Hunt. Development of Microsimulation Activity-Based Model for San Francisco: Destination and Mode Choice Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1777*, TRB, National Research Council, Washington D.C., 2001, pp. 25-35.

3. Algers, S., A.J. Daly, and S. Widlert. Modelling Travel Behaviour to Support Policy Making in Stockholm, In *Understanding Travel Behaviour in an Era of Change* (P. Stopher and M. Lee-Gosselin, eds.), Pergamon, Oxford, 1997.

4 Beser, M. and S. Algers. SAMPERS - The New Swedish National Travel Demand Forecasting Tool. In *National Transport Models: Recent Developments and Prospects* (L. Lundqvist and L-G. Mattsson, eds.), Springer, Stockholm. 2002.

5. Lerman, S.R., and C. Manski. On the Use of Simulated Frequencies to Approximate Choice Probabilities. In *Structural Analysis of Discrete Data and Econometric Applications* (C.F. Manski and D. McFadden, eds), MIT Press, Cambridge, MA, 1981, pp. 305-319.

6. Miller, E.J., M.J. Roorda, and J.A. Carrasco. A Tour-Based Model of Travel Mode Choice, *Transportation*, Vol. 32, No. 4, 2005, pp. 399-422.

7. Data Management Group. *Transportation Tomorrow Survey 1996: Data Validation*. Joint Program in Transportation, University of Toronto, Toronto, 1996.

8. Data Management Group. *TTS Version 3: Data Guide*. Joint Program in Transportation, University of Toronto, Toronto, 1997.

9. Ortúzar, J. de D., and L. Willumsen. *Modelling Transport*. Wiley, Chichester, NY, 2001.

10. Back, T. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms.* Oxford University Press, New York, 1996.

11. Kruchten, N. *Galapagos: A Distributed Parallel Evolutionary Algorithm Development Platform*. Bachelor's Thesis, University of Toronto, 2003.

12. Abdulhai, B., N. Kruchten, D. de Koning, and L. Kattan. Galapagos: Development Platform for Distributed Parallel Genetic Algorithms for Computationally Demanding ITS Optimization Problems. In *Proceedings of the 85th Annual Meeting of the Transportation Research Board*, CD-ROM. Transportation Research Board, National Research Council, Washington D.C., 2005.