

Development of an Estimation Procedure for an Activity-Based Travel Demand Model

W. Recker*, J. Duan & H. Wang

Department of Civil and Environmental Engineering and Institute of Transportation Studies,
University of California, Irvine, CA 92697

Abstract: *In this article, we implement an estimation procedure for a particular mathematical programming activity-based model to estimate the relative importance of factors associated with spatial and temporal interrelationships among the out-of-home activities that motivate a household's need or desire to travel. The method uses a genetic algorithm to estimate coefficient values of the utility function, based on a particular multidimensional sequence alignment method to deal with the nominal, discrete, attributes of the activity/travel pattern (e.g., which household member performs which activity, which vehicle is used, sequencing of activities), and a time sequence alignment method to handle temporal attributes of the activity pattern (e.g., starting and ending time of each activity and/or travel). The estimation procedure is tested on data drawn from a well-known activity/travel survey.*

1 INTRODUCTION

An essential element in the development of conventional trip-based disaggregate travel demand models is inferring the relative weights associated with potential components of the utility function that are determinants to a population's revealed selection of the decision variables (in the model estimation phase) with subsequent forecasts made using these weights in the application of the model. This particular aspect of the modeling process application of the activity-based research approach has remained a challenge, primarily due both to its complexity as well as the nature of solutions to the estimation problem. In contrast to conventional trip-based

disaggregate travel demand models in which the object of the estimation typically is the maximization of the likelihood of observing the outcome of a relatively simple choice (e.g., selection of a travel mode, travel route, or travel destination) or combinations of such choices (e.g., travel mode and route) from among a finite (countable) set of alternatives, activity-based models typically comprise a multitude of interrelated choices (e.g., travel mode-time of departure-participating actor pairings) for a sequence of activities (e.g., as shown in Figure 1) that are drawn from an infinite set of alternatives. And, although a utility maximization representation that bears resemblance to those found in discrete choice models can also be specified for an activity-based formulation, such as that described here, generally it has some notable exceptions that greatly complicate its application in empirical demand analysis—(1) the set of feasible solutions (alternatives) in the choice set is infinite, whereas that for standard discrete choice models is countable (and, usually small), (2) the solution vector comprises continuous, as well as discrete, variables, (3) although the overall solution represents a mutually exclusive choice, the solution vector itself is composed of components that are not generally mutually exclusive, (4) the components of the utility function are not directly interpretable as utility weights of attributes, but rather are related to these weights through a transformation matrix, and (5) the complexity of the constraint space generally precludes the type of closed-form probability result achievable with standard discrete choice models.

Although it was possible to generate the probability distributions of optimal solutions to the standard discrete choice problem either directly or through simple Monte Carlo simulation of the error terms (coupled

*To whom correspondence should be addressed. E-mail: wwrecker@uci.edu.

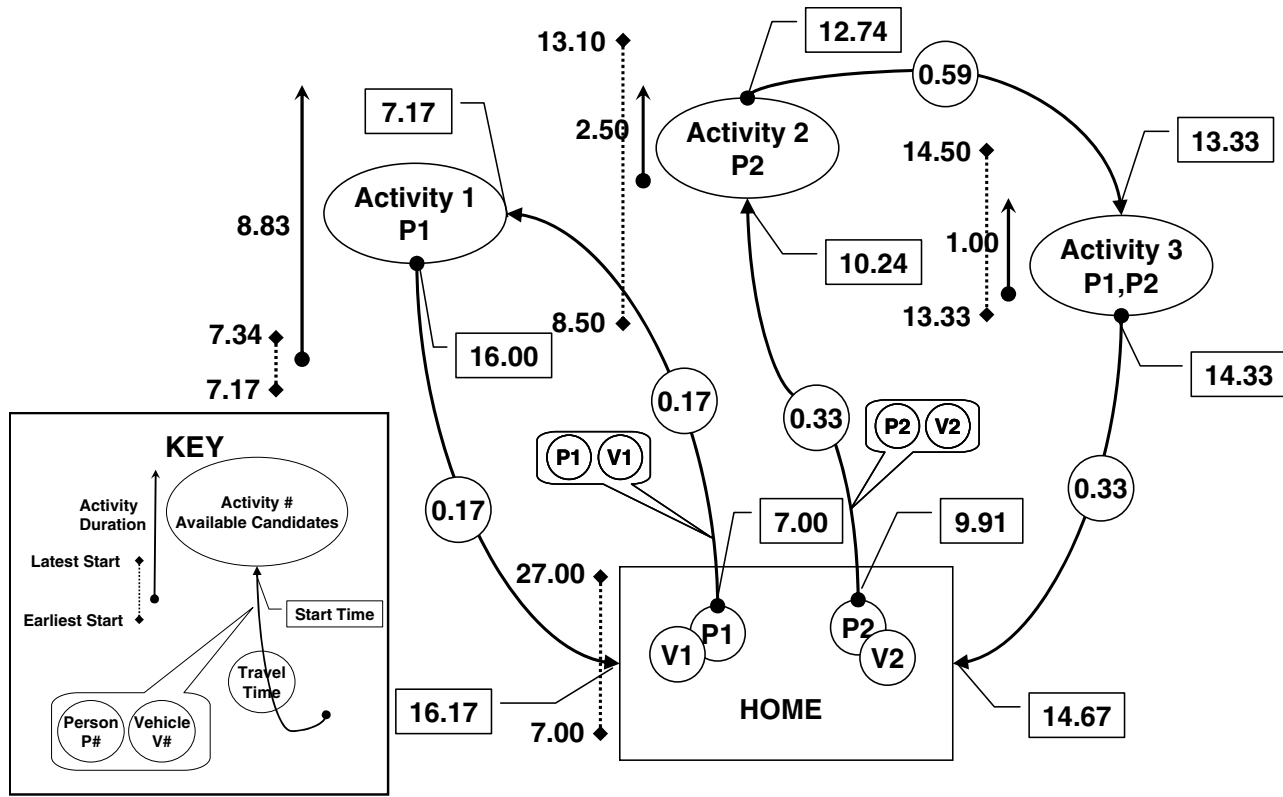


Fig. 1. Observed activity pattern for household 1.

to any standard integer programming technique), in the case of activity-based models this direct approach clearly is not possible. Rather, we are left with only approximate schemes to achieve this result. A second, and perhaps thornier, problem is that the solution vector contains continuous (i.e., the starting times of activities), as well as discrete, variables (i.e., person and vehicle assignments); hence, an infinity of alternatives exist. As such, the probability of any solution matching exactly the observed behavior is infinitesimally small. In dealing with continuous distributions, it is more proper to address the probability that an outcome lies within some error of a particular value or, alternatively, specify the continuous variables that define the choice alternative in terms of a “band” around the actual value; for example, an activity start time of t specified as a start time between $t - \delta$ and $t + \delta$.

In this article, we propose and implement an estimation procedure for the household activity pattern problem (HAPP) model (a mathematical programming activity-based model offered by Recker, 1995) to estimate the relative importance of factors associated with the spatial and temporal interrelationships among the out-of-home activities that motivate a household’s need or desire to travel. The procedure provides both the

necessary constraint considerations on the household’s decision alternatives within a utility-maximizing structure as well as a convenient mechanism for generating the set of feasible alternatives that are likely to be considered.

The HAPP model is in the form of a mixed integer linear programming model (MILP), that is, one comprising continuous variables (such temporal attributes of an activity pattern as the starting times of the associated activities) as well as discrete variables (e.g., attributes associated with the sequencing of activities, travel modes used, and persons performing the activities). Because of this complexity, an estimation approach based on a heuristic algorithm is the best (and, arguably, the only) available option for solution. From the collection of such heuristic algorithms, a genetic algorithm (GA) approach (Holland, 1992a,b) was used because of its considerable advantages for this particular application. To deal with the comparison of discrete attributes between the actual activity pattern and the predicted activity pattern, we use a variation of the multidimensional sequence alignment method (Arentze et al., 2002). The comparison of continuous attributes is based on a time sequence method that was developed as part of this research.

The article is organized as follows. We first provide a brief overview of the basic HAPP model as background for the current effort. We next state the general abstract form of the HAPP model and specify the activity-based HAPP model estimation problem. Then, we propose an estimation procedure using GA. We describe in detail how to implement the proposed estimation procedure using the GA and how to compare discrete and continuous attributes between the actual activity pattern and the predicted activity pattern. We use data drawn from a well-known activity-travel survey to test the implementation of the proposed estimation procedure, and then list further research needed to extend the estimation procedure.

2 GENERAL FORM OF THE HAPP MODEL AND ITS ESTIMATION PROBLEM

The general form of the HAPP mathematical program formulation of the travel/activity decisions for a particular household, say, i , during some time period is represented by:

$$\begin{aligned} & \text{Minimize } \mathbf{Z}(\mathbf{X}_i) = \mathbf{B}'_i \cdot \mathbf{Y}(\mathbf{X}_i) \\ & \text{Subject to} \\ & \mathbf{A}\mathbf{X}_i \leq \mathbf{0} \end{aligned} \quad (1)$$

where

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{T} \\ \mathbf{X}^v \\ \mathbf{H} \end{bmatrix}, \quad \mathbf{T} = [T_u \geq 0], \quad \mathbf{X}^v = \left[X_{uw}^v = \begin{cases} 0 \\ 1 \end{cases} \right],$$

$$\mathbf{H} = \left[H_{uw}^\alpha = \begin{cases} 0 \\ 1 \end{cases} \right]$$

The output \mathbf{X}_i of the optimization of each household i are specified by the following decision variables:

- T_u : the time at which participation in activity u begins
- X_{uv}^v : binary decision variable equal to unity if vehicle v travels from activity u to activity w , and zero otherwise
- H_{uw}^α : binary decision variable equal to unity if household member α travels from activity u to activity w , and zero otherwise.

In the above, $\mathbf{Y}(\mathbf{X}_i)$ denotes a vector of partwise utility components constructed from the decision variables, and \mathbf{B}_i is a vector of coefficients that defines the relative contributions of each of the partwise utility components

to the overall disutility of the travel regime to the household. Descriptively, the constraint sets $\mathbf{A} \cdot \mathbf{X}_i \leq \mathbf{0}$ for this MILP are classified into six groups: (a) routing constraints that define the allowable spatial movement of vehicles and household members in completing the household's activity agenda; (b) scheduling constraints specify the relationship of arrival time, activity begin time, and waiting time, and continuity condition along the temporal dimension; (c) assignment constraints that are applied to match the relations between activity participation and vehicle usage as well as activity performers (household members); (d) time window constraints that are used to specify available schedules for activity participation; (e) coupling constraints that define the relations between vehicle-related variables and member-related variables; and (f) side constraints including budget, capacity and rules for ride-sharing behavior. With the exception of the side constraints (i.e., classification "F" above), these constraints capture the physical conditions that ensure that each member of the household, as well as each vehicle used by the household, have a consistent, continuous, path through time-space that results in all of the activities on the household's agenda being successfully completed. The reader interested in a detailed derivation and explanation of these constraints is referred to the original work by Recker (1995).

The solution vector, \mathbf{X}_i^* to Equation (1) represents the household's utility-maximizing behavior, relative to the prescribed objective $Z(\mathbf{X}_i)$, with regard to completing its activity agenda. The solution patterns reveal personal travel behavior and activity participation within a household context, while preserving the concept that the need for travel originates from participation in activities, that travel constitutes the linkage between activities, and in which all of the required components are contained in the activity scheduling problem.

The methodology described above has been applied successfully to a number of transportation applications to explore issues relating to such areas as vehicle emissions, accessibility, trip-chaining, ride-sharing, and travel time reduction (see, e.g., Recker and Parimi, 1999; Recker et al., 2001). However, in all of these applications, the specification of the objective function is *prescribed* by the analyst; for example, the minimization of emissions produced by travel, rather than *estimated* from the revealed choices made by members of the household. The inference of the relative weights associated with potential components of the utility function that are determinants to a population's revealed selection of the decision variables is necessary before conventional application of the model to forecasts can

be made; this particular aspect of the application of the activity-based research approach has remained a challenge for the reasons mentioned above.

As common to any statistical inference, the analyst cannot directly observe \mathbf{B}_i ; rather, an estimate, $\hat{\mathbf{B}}_i$, is sought that can be inferred from the observed behavior, $\tilde{\mathbf{X}}_i^*$. The goal, then, is to find the $\hat{\mathbf{B}}_i$ that minimizes some prescribed error ε_i between the solution vector \mathbf{X}_i^* and the observed behavior $\tilde{\mathbf{X}}_i^*$. In the case of discrete choice analysis, this is usually accomplished with maximum likelihood estimation of $\hat{\mathbf{B}}_i \equiv \hat{\mathbf{B}}, \forall i$, that is, assuming that the utility weights are common across observations, and assuming that the error terms are independently identically distributed (IID). The standard application of maximum likelihood requires that the model choice probabilities be a differentiable function of the parameters contained in \mathbf{B} . Without such stipulation, the maximization process could be accomplished using some heuristic search. In the case of the HAPP model (i.e., Equation (1)), the only option available is a heuristic. The particular form of the HAPP model lends the accompanying estimation problem to solution by GA (Rudolph, 1994). Owing to the specification of \mathbf{X}_i in Equation (1), we define a three-tuple

$$\mu(\hat{\mathbf{B}}_i^p) = \{\mu_1(\hat{\mathbf{B}}_i^p), \mu_2(\hat{\mathbf{B}}_i^p), \mu_3(\hat{\mathbf{B}}_i^p)\} \quad (2)$$

where its elements,

$$\mu_1(\hat{\mathbf{B}}_i^p) = \text{err}|\mathbf{T}_i(\hat{\mathbf{B}}_i^p) : \tilde{\mathbf{T}}_i | \tilde{\mathbf{X}}_i^y, \tilde{\mathbf{H}}_i |$$

$$\mu_2(\hat{\mathbf{B}}_i^p) = \text{err}|\mathbf{X}_i^y(\hat{\mathbf{B}}_i^p) : \tilde{\mathbf{X}}_i^y | \tilde{\mathbf{H}}_i, \tilde{\mathbf{T}}_i |,$$

$$\mu_3(\hat{\mathbf{B}}_i^p) = \text{err}|\mathbf{H}_i(\hat{\mathbf{B}}_i^p) : \tilde{\mathbf{H}}_i | \tilde{\mathbf{X}}_i^y, \tilde{\mathbf{T}}_i |$$

represent the error between the observed and predicted values of the respective components of the particular household observation's activity pattern, for any particular candidate solution $\hat{\mathbf{B}}_i^p, p = 1, \dots, n_p$; here, the particular norm used to define $\text{err}|\cdot|$ depends on the character of the associated variable, for example, for the continuous time variables \mathbf{T}_i a Euclidean norm is appropriate, whereas for the discrete variables \mathbf{X}_i^y and \mathbf{H}_i metrics based on Levenshtein distance are appropriate. We interpret $\mu(\hat{\mathbf{B}}_i^p)$ as a multi-objective measure that indicates the fitness of any potential solution vector, $\hat{\mathbf{B}}_i^p$, in minimizing the discrepancies between the observed and model values. Then, using standard procedures in genetic algorithms, we generate an initial population of candidate solutions $\hat{\mathbf{B}}_i^p, p^1 = 1, \dots, n_p^1$, and perform mating, crossover, and mutations (Holland, 1992a; Michalewicz, 1994) according to predescribed probability rules based on respective fitness scores to produce succeeding populations of candidate solutions, $\hat{\mathbf{B}}_i^p, p^j = 1, \dots, n_p^j; j = 1, \dots, J$, until the population converges to within sufficient tolerance of the observed

values; that is, $\hat{\mathbf{B}}_i = \hat{\mathbf{B}}_i^{p^j} \ni \lim_{j \rightarrow \infty} |\mu(\hat{\mathbf{B}}_i^{p^j})| \leq \varepsilon^*$, where ε^* is the prescribed error tolerance.

Q1

3 IMPLEMENTATION OF THE ESTIMATION PROCEDURE

The estimation procedure using the GA described in the previous section was applied with one minor modification. To save computation time, we compressed the three-tuple into a two-tuple in the objective function by combining $\mu_2(\hat{\mathbf{B}}_i^p)$ for \mathbf{X}_{uw}^v and $\mu_3(\hat{\mathbf{B}}_i^p)$ for \mathbf{H}_{uw}^α into a single measure $\mu_{23}(\hat{\mathbf{B}}_i^p)$, which we refer to as the "sequence distance" between predicted activity pattern (PAP) and observed activity pattern (OAP), and is measured by the number of steps needed to equalize the nominal (discrete) aspects of the two sequences, related only to such qualitative properties of activity pattern as which activity to do at any particular point in the sequence of activity participation, which vehicle to use and, which person to be assigned to the activity. This is nonrestrictive because both \mathbf{X}_{uw}^v and \mathbf{H}_{uw}^α in the general form of the HAPP model are (nominal) qualitative variables and, in the demonstration application reported here, weights on each of the three dimensions were arbitrary and assumed to be equal. We calculate sequence distance using the multi-dimensional sequence alignment method (MDSAM) (Arentze et al., 2002) to compare multiple qualitative properties, \mathbf{X}_{uw}^v and \mathbf{H}_{uw}^α , under this common measure. The MDSAM process optimally (i.e., minimum effort) "re-constructs" the nominal parameters (i.e., integer variables of the HAPP model) of the predicted household activity sequence to perfectly align with the observed sequence of the particular household (i.e., to match the observed assignments of household member and vehicle to each activity in the household's activity plan, as well as the observed sequence in which the respective activities were performed). This process calculates the minimal effort required, $\mu_{23}^i(\hat{\mathbf{B}}_i^p)$, to perfectly align the two activity sequences, as follows:

1. Calculate the Levenshtein distance between each pair of unidimensional sequences in the source and target multidimensional sequences, using the unidimensional sequence alignment method (UDSAM) of MDSAM.
2. Construct a *tribranch optimum trajectory tree* for each pair of unidimensional sequences, according to Levenshtein distances above.
3. Search the trees to find all optimum trajectories (paths) for each pair of unidimensional sequences and record the operation set of each trajectory. Delete, substitute, and insert operations

are recorded, but identity operations are ignored because they are cost-free. All trajectories are stored in a dynamic vector.

4. Produce all possible element-operation-based combinations from all optimum trajectories of every pair of sequences, and then rearrange each element-operation-based combination to segment-operation-based combination. To calculate more efficiently, hash tables are used to store all possible segments (no duplicate) in each trajectory.
5. Calculate the costs of each segment-operation-based combination above.
6. Compare all costs, and get the optimum combinations with the lowest cost.

Following this stage in the process, the PAP of any particular household in the data set will be sequentially aligned with the OAP, but the PAP's corresponding activity start times will generally be different from those observed— $\mu_1(\hat{\mathbf{B}}_i^p)$ is used as a metric for this aspect of the error between the OAP and PAP. For ease of calculation, we define $\mu_1(\hat{\mathbf{B}}_i^p)$ as the “temporal similarity” between PAP and OAP, rather than as a dissimilarity measure, and simply negate this term in the overall measure of the error between the OAP and PAP. Temporal similarity is related to such continuously varying properties of the activity pattern as starting time and ending time of each particular activity and/or travel that comprises the activity pattern. In application, $\mu_1(\hat{\mathbf{B}}_i^p)$ indicates how similar the two activity sequences are in their temporal dimension, that is, the similarity \mathbf{S} (overlap time) between identical activities in the two activity sequences. This second stage has two distinct components: (1) global temporal shift of the entire pattern, and (2) local temporal comparison of each activity. We compute this temporal similarity as follows:

1. Readjust the predicted sequence according to the observed sequence performed by the same person, that is, reshuffle the predicted activity sequence to the same order as the observed sequence. During the reshuffle, if an activity in the observed sequence is missing in the predicted sequence, then insert this activity into predicted sequence, but with start time = -1 . If there is an activity in the predicted, but missing in observed sequence, simply delete it from predicted sequence.
2. Using the predicted activity sequence in step 1 as a base, produce the corresponding starting time and ending time for each activity in the predicted and observed patterns. Activities in the two lists have a one-to-one relationship relative to the pair (Activity, Person).

3. Global alignment to decide the starting point of local comparison in the two lists.

Do Loop

{

Move pointers forward in two lists one by one;

If (two activities in current pointer position are same)

{

If (both of start times > 0)

Do local time sequence comparison;

}

}while (all activities are scanned)

Return optimal result based on a given dimension (see below)

4. Local time comparison. For each dimension (e.g., person performing, vehicle mode in use, etc.), calculate the time overlap of the predicted activity pattern with its corresponding observed activity pattern—the common time in this comparison is a measure of the time similarity for this particular pair of activities. Return to the global alignment (step 3) to continuously compare other pairs of activities in the activity sequence.

With the consolidation of error in all of the nominal/discrete aspects of the activity pattern collapsed into a single measure, $\mu_{23}^i(\hat{\mathbf{B}}^{p_i})$, the three-tuple error metric in Equation (2) becomes $\mu(\hat{\mathbf{B}}_i^p) = \{\mu_1(\hat{\mathbf{B}}_i^p), \mu_{23}(\hat{\mathbf{B}}_i^p)\}$. In application, the goal of minimizing the error between the OAP and PAP relative to this multidimensional metric ultimately relies on assigning relative importances to the metric's components to produce a scalar quantity. Here, we accomplish this with the following transformation:

$$\mu(\hat{\mathbf{B}}_i^p) = \{\mu_1(\hat{\mathbf{B}}_i^p), \mu_{23}(\hat{\mathbf{B}}_i^p)\} \rightarrow \mu(\hat{\mathbf{B}}_i^p) = a\mu_{23}(\hat{\mathbf{B}}_i^p) - b\mu_1(\hat{\mathbf{B}}_i^p) + c; a, b, c \text{ positive constants}, \quad (3)$$

examples of which are commonly referred to in the literature as a form of “edit distance” measure. The objective then of the PAP of any particular household matching as closely as possible its OAP is simply stated as $\text{Min}_{\hat{\mathbf{B}}_i^p} \mu(\hat{\mathbf{B}}_i^p)$; that is, select the utility weights comprising $\hat{\mathbf{B}}_i^p$ that minimize the weighted sum of the number of steps required to align the respective activity sequences (PAP and OAP), that is, actors performing the activities, and vehicles used to access the activities, while maximizing the temporal overlap between the two patterns. The constant c is used to scale the objective function to some range, such as $[0, 1]$.

For a sample of N households, each household has its own PAP and OAP; correspondingly, each pair of PAP

and OAP will have its own “error” for any particular $\hat{\mathbf{B}}_i^p$. In applying Equation (3) as the kernel for estimating the optimal common utility weights (in the sense of minimizing the total error between the respective OAPs and PAPs for the sample), that is, $\hat{\mathbf{B}}_i \equiv \hat{\mathbf{B}}, \forall i$, for the sample of N households, we form the objective function for determining $\hat{\mathbf{B}}$ as

$$\begin{aligned} \text{Min}_{\hat{\mathbf{B}}} \mu(\hat{\mathbf{B}}) &= \sum_{i=1}^N w_i \mu(\hat{\mathbf{B}}_i) \\ &= \lim_{j \rightarrow \infty} \sum_{i=1}^N w_i (a \mu_{23}^i(\hat{\mathbf{B}}_i^{p^j}) - b \mu_1^i(\hat{\mathbf{B}}_i^{p^j}) + c) \end{aligned} \quad (4)$$

where w_i is a weighting factor that can be assigned to account for differential importance placed on the error contribution of household i to the total error,¹ a , b , and c are positive constants, and where $\mu_{23}^i(\hat{\mathbf{B}}_i^{p^j})$ and $\mu_1^i(\hat{\mathbf{B}}_i^{p^j})$ are scaled to $[0, 1]$. The selection of values for constants a and b is discretionary and reflect the relative importance of errors in the corresponding aspects to the study at hand. For example, “time utilization” studies might typically place greater emphasis on calibrating toward the correct assignment of individuals to activities (i.e., $|a| > |b|$), whereas in a typical transportation planning application, the emphasis more likely would be on the timing of trips (i.e., $|a| < |b|$).

The steps in the estimation are as follows:

1. Generate a population, say the j th, of n_p^j candidate solutions $\hat{\mathbf{B}}_i^{p^j} = \hat{\mathbf{B}}^{p^j}, \forall i, p^j = 1, \dots, n_p^j$.

Because the signs of the coefficients of $\hat{\mathbf{B}}_i^{p^j}$ typically are known, the utility components can always be specified in such a manner that the elements of $\hat{\mathbf{B}}_i^{p^j}$ are strictly nonnegative; and, because the solution is unaffected by the scaling of the utility coefficients, they can be scaled to be in a given range. We first encode each coefficient in $\hat{\mathbf{B}}_i^{p^j}$ as a binary (0, 1) string. Then, all of these strings are concatenated together in sequence to form a chromosome; the elements of each chromosome are referred to as genes.

We encode the integer part and decimal fraction of each coefficient separately. For the integer part, we define a range of $[0, 2^m]$, where m is the length of the binary string for the integer part. For the decimal fraction, we define a range of $[0, 2^n]$, where n is the length of the binary string for decimal fraction. So, the first m bits of an $m + n$ binary string represent the integer part of a coefficient and its last n bits represent the decimal fraction of the coefficient. Decoding is

accomplished by separating the chromosome into two parts. The first part (length = m) is for integer part, which can be transformed directly from binary string to decimal; the second part (length = n) is for decimal fraction, which is transformed by a two-step process: (1) transform the binary string to decimal integer and (2) divide the decimal integer obtained from the first step by 2^n .

2. Each chromosome, $\hat{\mathbf{B}}^{p^j}, p^j = 1, \dots, n_p^j$ in the population is assigned a “fitness score”

$$\begin{aligned} \mu(\hat{\mathbf{B}}^{p^j}) &= \sum_{i=1}^N w_i \mu(\hat{\mathbf{B}}_i^{p^j}) = \sum_{i=1}^N w_i (a \mu_{23}^i(\hat{\mathbf{B}}_i^{p^j}) \\ &\quad - b \mu_1^i(\hat{\mathbf{B}}_i^{p^j}) + c); p^j = 1, \dots, n_p^j \end{aligned} \quad (5)$$

Here, for demonstration purposes, both $\mu_{23}^i(\hat{\mathbf{B}}_i^{p^j})$ and $\mu_1^i(\hat{\mathbf{B}}_i^{p^j})$ are scaled to $[0, 1]$. To scale $\mu_{23}^i(\hat{\mathbf{B}}_i^{p^j})$ to $[0, 1]$, the maximum possible value of the sequence distance, as computed by the MDSAM method, is required. The results reported here are based on the assumption that insertion and deletion costs are 1, whereas the substitution cost is 2 (effectively a deletion followed by an insertion). We also assume that each discrete attribute may have its own relative cost weight; we label the largest cost weight among the attributes as ω_{\max} . If, in addition, the OAP and PAP sequences have different lengths m and n ($m > n$), the maximal MDSAM operation cost will be $\omega_{\max} \cdot [2n + (m - n)] = \omega_{\max} \cdot (m + n)$. Therefore, the MDSAM result scaled to $[0, 1]$ could be obtained through normalization by $\omega_{\max}^*(m + n)$. Correspondingly, the normalization value for $\mu_1^i(\hat{\mathbf{B}}_i^{p^j})$ will be the maximum time coverage, or the sum of all the durations in the observed sequence.

Scaled as above, by selecting $(a, b, c) = (1/2, 1/2, 1/2)$, the sequence distance and temporal overlap results can be combined to represent the integrated “error” between the predicted and the observed sequences for any given household i as

$$\begin{aligned} \mu(\hat{\mathbf{B}}_i^{p^j}) &= a \mu_{23}^i(\hat{\mathbf{B}}_i^{p^j}) - b \mu_1^i(\hat{\mathbf{B}}_i^{p^j}) + c \\ &= \frac{(\text{Sequence Distance} - \text{Temporal Overlap}) + 1}{2} \end{aligned} \quad (6)$$

Because both $\mu_{23}^i(\hat{\mathbf{B}}_i^{p^j})$ and $\mu_1^i(\hat{\mathbf{B}}_i^{p^j})$ are scaled to $[0, 1]$, the range of this result will be $[0, 1]$.

3. Each chromosome in population j is assigned a series of three “probability of reproduction” values, each of which is inversely proportional to its fitness $\mu(\hat{\mathbf{B}}^j)$ relative to the other chromosomes in the population, that is, the lower the value of the fitness measure (distance to the observed), the higher the probability of reproduction.
4. According to the assigned probabilities of reproduction, a new population, $j + 1$, is generated by “mating” a selection drawn according to the probability.
5. The selected pairs of chromosomes generate offspring via the use of such specific genetic operators as crossover and gene mutation. Crossover is applied to two chromosomes (parents) and creates two new chromosomes (offspring) by selecting a random position along the coding and then splicing the section that appears before the selected position in the first string with the section that appears after selected position in the second string, and vice versa.
6. The process is halted if a suitable solution is found. Otherwise, the process returns to step 2, where the new chromosomes are scored and the procedure iterates.

To initiate the GA procedure, we generate an initial population randomly. Subsequently, new populations are generated in the mating pool using the GA operations described above, based on the given probabilities.

For the results presented here, the HAPP mixed-integer linear programming model is solved using the CPLEX algorithm in the GAMS software package developed by the World Bank. Under the current implementation, the algorithm takes as input the composition (e.g., number of household members, number of vehicles by type), activity plan (e.g., number, location, and type of activities to be completed), and constraints (e.g., activity time windows, preassigned roles) faced by a particular household, and a set of model coefficients (i.e., candidate utility weights), which are decoded from the chromosome produced by the GA (each chromosome includes all coefficients in the HAPP model, represented as a binary string). Based on this information, the HAPP model outputs the predicted activity pattern that maximizes the objective utility function for the particular utility weights.

4 INITIAL TESTS

To demonstrate the estimation procedure, we randomly selected household activity patterns of 65 two-

adult-member households extracted from the South-west Washington and Oregon Area 1994 Activity and Travel Behavior Survey (NuStats, 1995; Cambridge Systematics, 1996), which contains sufficiently detailed information, including comprehensive travel/activity diaries (with mode availability) and regional transportation network model, to support an application of the HAPP model used in this study. The activities considered in the analysis are all out-of-home activities and in-home meals for the first day of the two-day activity diary; in-home activities other than meals are assumed to be discretionary and flexible. The mean number of activities per household is 6.56 with a standard deviation of 2.23; the mean number of activities performed by all licensed drivers within a household is 5.84 with a standard deviation of 1.87. About 62% of the activities were out-of-home activities that required travel (17.3% work, 14.6% general shopping, and 23.9% from the remaining categories), with an average travel time per activity requiring travel of approximately 18 minutes. The mean total travel time of those individuals in the household who traveled is 0.95 hours. The mean duration of a meals activity is 1 hour and the corresponding mean duration of work and shopping activities is 6 hours and 0.8 hours, respectively. Approximately 75% of the households have two vehicles and about 93% of the households have more than one vehicle; 90% of the households had two licensed drivers, whereas 8% had only one licensed driver.

Because the HAPP model represents activity/travel behavior of the collective members of a household, detailed information is required on travel and activity participation for each member, as well as transportation supply information (including household vehicle holdings and network travel times). To help define the time window constraint space, the average activity starting and ending times for each activity type were computed for the whole sample to provide benchmark information on the temporal flexibility of the activities. Shortest path travel times between all activity locations of each household have been generated for all the households in the sample using TRANSCAD. This procedure allows for the exploration of all possible activity/travel linkages for each household, which is fundamental to the optimization procedure.

Following Equation (1), for each of the households drawn from the sample for this initial test, a common disutility function was assumed of the form:

$$\text{Minimize } \mathbf{Z}(\mathbf{X}_i) = \mathbf{B}'_i \mathbf{Y}(\mathbf{X}_i) \quad (6)$$

where \mathbf{B}_i is a vector of unknown parameters (to be estimated), and where

$$\mathbf{Y}(\mathbf{X}_i) = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} \sum_{v \in \mathbf{V}} \sum_{u \in \mathbf{N}} \sum_{w \in \mathbf{N}} c_{uw}^v X_{uw}^v \\ \sum_{v \in \mathbf{V}} \sum_{u \in \mathbf{N}} \sum_{w \in \mathbf{N}} t_{uw}^v X_{uw}^v \\ \sum_{u \in \mathbf{P}^+} (T_{u+n} - T_u - s_u - t_{u,u+n}^v) \\ \sum_{u \in \mathbf{P}^+} (T_u + s_u - b_u) \\ \sum_{u \in \mathbf{P}^-} (T_u - b_u) \\ \sum_{v \in \mathbf{V}} (T_{2n+1}^v - T_0^u) \\ \sum_{v \in \mathbf{V}} \sum_{u \in \mathbf{P}^+} c_v X_{0u}^v \\ \sum_{u \in \mathbf{P}^+} \left(T_u - \sum_{v \in \mathbf{V}} \sum_{w \in \mathbf{N}} (T_w + t_{wu}^v) X_{wu}^v \right) \end{bmatrix} \quad (7)$$

The notation used in Equation (7) is the same as that in Recker (1995). Succinctly, c_{uw}^v and t_{uw}^v are, respectively, the travel cost and time for a trip from the location of activity u to the location of activity w by vehicle v ; s_u is the duration of activity u ; b_u is the close of the availability of activity u ; \mathbf{V} is the set of vehicles available to the household; \mathbf{P}^+ is the set of n out-of-home activities; \mathbf{P}^- is the set of n return-home activities; nodes 0 and $2n + 1$ are, respectively, the beginning- and end-of-day locations; and $\mathbf{N} = \{0, \mathbf{P}^+, \mathbf{P}^-, 2n + 1\}$. A qualitative interpretation of the variables comprising the assumed disutility function is provided in Table 1.

Table 2
Genetic algorithm parameters

| Genetic parameter | Value |
|-----------------------|---------------------|
| Chromosome length | 80 |
| Alphabet size | 2 |
| Population size | 20 |
| Stop criterion | 25 generations |
| Crossover probability | 1.0 |
| Mutation probability | 0.05 |
| Selection method | Roulette wheel |
| Crossover option | Two-point crossover |
| Constants a, b, c | 0.5, 0.5, 0.5 |

Based on this objective function, solutions to the corresponding HAPP model for each household were obtained for an initial population of 20 chromosomes, that is, $\hat{\mathbf{B}}_p$, $p = 1, \dots, 20$, each having eight genes corresponding to the eight unknown coefficients of y_i . The resulting fitness values were determined according to the two-stage process described above (i.e., minimizing the Levenshtein distance between the observed and model-predicted discrete sequence components, followed by the time sequence adjustment phase), and input to the GA. The parameters used by the GA are displayed in Table 2.

4.1 Demonstration application inferring \mathbf{B}_i for a single household

In this section, we provide some detailed results of application of the estimation procedure to a series of single, independent, households; that is, those for which we did not invoke the standard constraint of having common coefficients, $\hat{\mathbf{B}}_i^{p^j} = \hat{\mathbf{B}}^{p^j}$, $\forall i, p^j = 1, \dots, n_{p^j}$. Shown

Table 1
Qualitative interpretation of components of disutility function

| Variable | Interpretation |
|----------|--|
| y_1 | Total cost of travel to the household. |
| y_2 | Total travel time expended by members of the household. |
| y_3 | The difference between the time that a household member arrives home from a particular activity and the time he/she started the activity; a measure of delay time in returning home from an activity because he/she went to another activity prior to returning home (i.e., trip chained). |
| y_4 | A measure of how close a household member comes to being unable to complete an activity before its window closes, that is, the risk of not being able to perform an activity if some stochastic delay (e.g., congestion) had arisen. |
| y_5 | Same as y_4 above, but with respect to returning home from an activity no later than that needed/desired. |
| y_6 | Total extent of the travel day; that is, how spread out the travel is. |
| y_7 | The "cost" associated with using more vehicles than absolutely necessary; for example, using two different vehicles, rather than one. |
| y_8 | Total waiting time (at destination) before performing activities. |

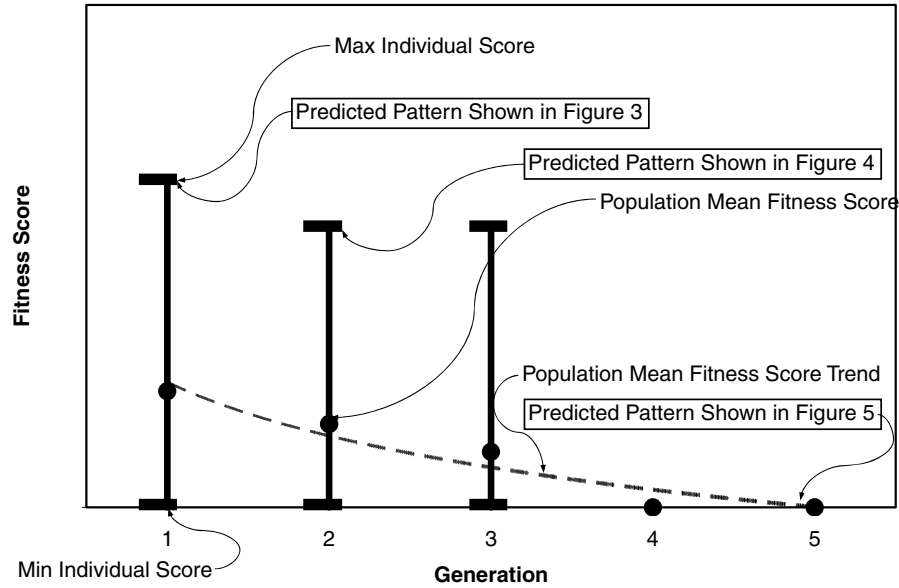


Fig. 2. Convergence of genetic algorithm for household observation 1.

below, in Figures 1, 7, and 9, are schematic representations of the actual, observed activity patterns of three households, each of which is typical of those in the data set that comprised two persons of driving age and having two vehicles available to the household.

Figure 2 shows the convergence of the GA for household observation 1; displayed are statistics on the overall fitness values for five generations of the population (Because the scale on the fitness axis is arbitrary, the numerical values are not displayed.). Because the con-

straints on the problem limit the number of feasible patterns, and feasibility is a necessary condition, even the initial population (as well as subsequent generations) contained one or more “solutions” that matched the observed pattern of Figure 1.

The algorithm converged to the actual, observed, pattern in only four generations. The predicted patterns associated with the worst (i.e., greatest positive) fitness values for generations 1 and 2 are displayed in Figures 3 and 4, respectively. In these figures, errors in temporal

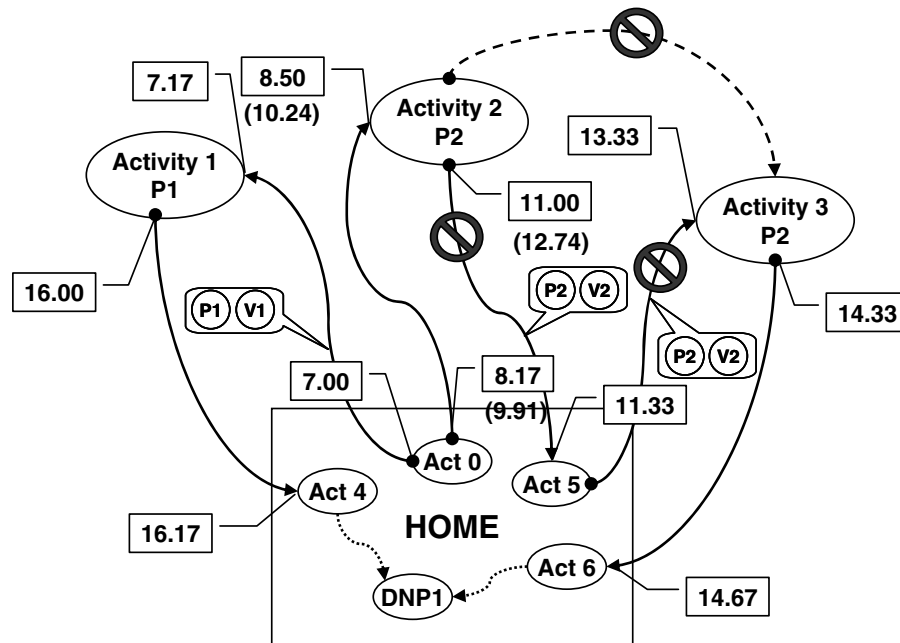


Fig. 3. HAPP-predicted pattern for household 1, “worst fitness” value: generation 1.

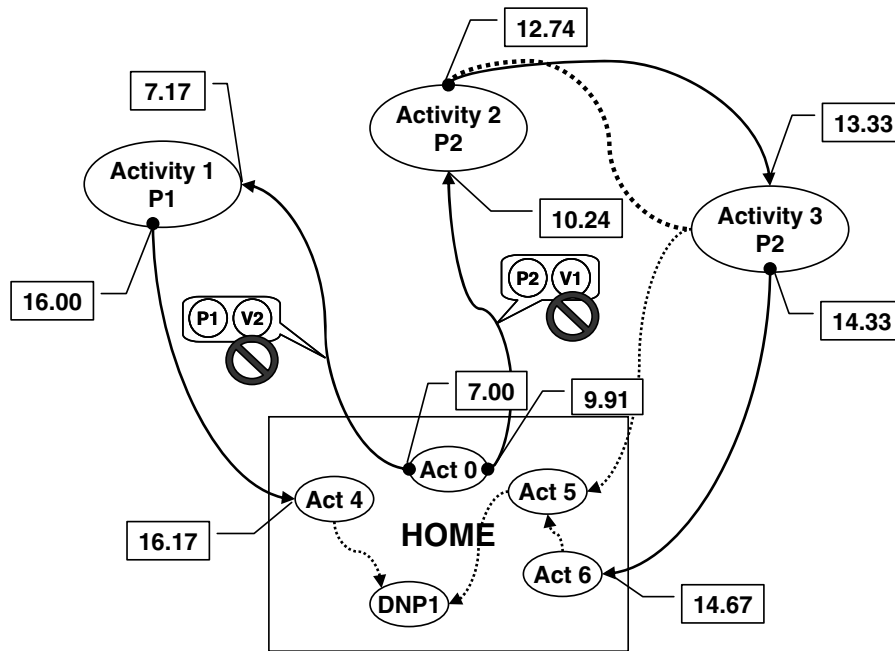


Fig. 4. HAPP-predicted pattern for household 1, “worst fitness” value: generation 2.

values (e.g., starting times) have been identified by placing the correct, observed, value in parentheses next to the predicted values; errors in discrete variables, such as paths and vehicles used, have been designated with the symbol \otimes . We note that the predicted patterns (i.e., as output by the HAPP model) have an expanded version of the “home” node (as required by the HAPP formulation) that includes the “virtual” activities corresponding to the initial departure from home (labeled “ACT 0”), the return-to-home activity for each real activity $i = 1, \dots, n$ on the household’s activity agenda (labeled “ACT $i + n$ ”), and the end-of-day activity (labeled “DNP1”). The observed pattern for observation 1 (Figure 1) indicates that person 1 departed from home in vehicle 1 at 7.00 hours (time is in military decimal), arriving at the destination to perform activity 1 at 7.17, where participation begins immediately upon arrival. Immediately following completion of the activity, person 1 returns home. Person 2 leaves home at 9.91 in vehicle 2, and trip-chains activities 2 and 3 before returning home. The predicted pattern in Figure 3, which corresponds to the HAPP pattern with the worst fitness score relative to the observed pattern, exhibits both activity linkage errors and temporal deviations from the observed (Figure 3). It fails to capture the trip chaining behavior of person 2, as well as not correctly matching the initial departure from home for person 2 and the start and end times for activity 2; this pattern is eliminated after the first generation by the GA.

The only discrepancy in the pattern with the worst fitness score in generations 2 and 3 is in the predicted vehicle used by persons 1 and 2 in completing their activities (Figure 4). By generations 4 and 5, all 10 of the predicted patterns in each of these generations matched the observed patterns exactly (Figure 5).

The corresponding convergence in the utility coefficients to produce an exact match between the predicted and observed patterns is displayed in Figure 6, where the mean value of the estimated coefficients, together with their standard deviation error band are plotted for each generation. (We note that, as with the case in traditional demand modeling, the set of model coefficients that replicates the observed outcome generally is not unique.) After five generations, most coefficients have converged to a single value for the entire population; those that have not differ from each other by relatively small amounts.

Shown in Figure 7 is the observed pattern for a second household, household 2. The activity pattern of this household included three specific “in-home” activities (meals): Activity 1 (person 1’s breakfast, and activities labeled 3 and 5 (joint dinner)). Shown in Figure 8 is the corresponding predicted pattern with the best fitness score after five generations. By the fifth generation the estimated utility coefficients produced by the GA resulted in a prediction that varied from the observed only in the timing of activity 4; activity 4 is predicted to commence 0.17 hours later than observed, and following its completion the prediction has person 2 waiting at

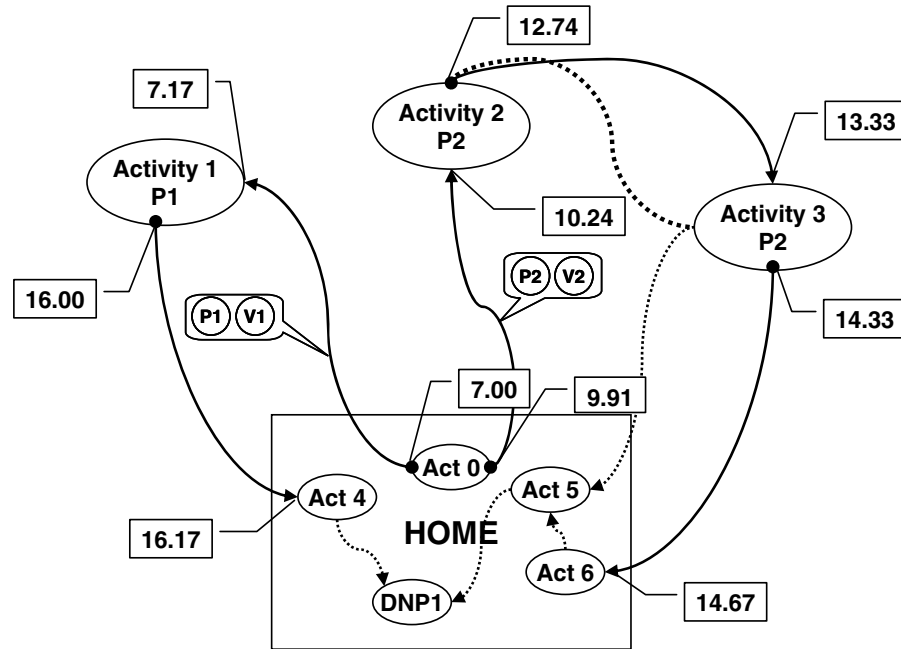


Fig. 5. HAPP-predicted pattern for household 1, generations 4 and 5.

the location of activity 4 for 1.5 hours before returning home to participate in activity 4. (There is no “penalty” in the specified utility function for wait time incurred at the location of the previous activity; only for wait time incurred at the location of the current activity, prior to its commencement.)

The final household considered in this initial test (household 3) is distinct from the previous two in that the observed pattern, as reported in the survey, contained reporting errors that rendered the reported pattern infeasible (Figure 9). Specifically, the pattern for this household includes carpooling to a joint out-of-home activity (labeled activity 1 and 2 for persons 1 and 2, respectively) that has been reported incorrectly in the survey. Although using the same vehicle (vehicle 1) and arriving at the same destination, person 2 reports an activity start time and return home time that differ from that reported by person 1 by 0.12 hours. After five generations, the GA failed to identify a set of coefficients that would result in the HAPP model capturing the observed (albeit misreported) carpooling behavior (Figure 10). Nonetheless, the model was able to replicate the reported times for activities 1 and 2 exactly; the misclassification of the person and vehicle assigned to these two activities is a semantic difference because the two activities are, in reality, the same. The only other difference between the “best” prediction and the observed is a 0.29-hour difference in the schedule times associated with activity 3, and the vehicle used to complete that activity.

4.2 Results of estimation of utility coefficients on a sample of 65 households

In this section, we give summary results of application of the estimation procedure to the sample of 65 randomly selected two-member households under the usual assumption of homogeneity, that is, $\hat{\mathbf{B}}_i = \hat{\mathbf{B}}, \forall i$. The results are based on 25 generations of populations of size 20, using the GA described in the previous sections, and based on GA parameters shown in Table 2. Because there are no well-defined theoretical distributional properties related to the error of the estimation, we rely on several “goodness-of-fit” indicators to capture the efficacy of the estimation: (1) the mean edit distance $\mu(\hat{\mathbf{B}})$ [0, 1] for the sample, (2) the percentage of activity sequence orderings predicted correctly, and (3) the total difference between the starting times for all activities in a household member’s observed and predicted activity patterns. Shown in Figure 11 are the distributions of these measures across the entire population covered by the random initial population and subsequent 25 generations (i.e., based on $26 \times 20 = 520$ sets of parameters, $\hat{\mathbf{B}}$).

These results tend to indicate that, for the sample considered, constraints play a dominant role in the execution of the observed activity patterns—relatively high agreement between the observed and predicted patterns is achieved over a broad range of ascribed utility weights represented by the components of $\hat{\mathbf{B}}$. This observation regarding the apparent dominance

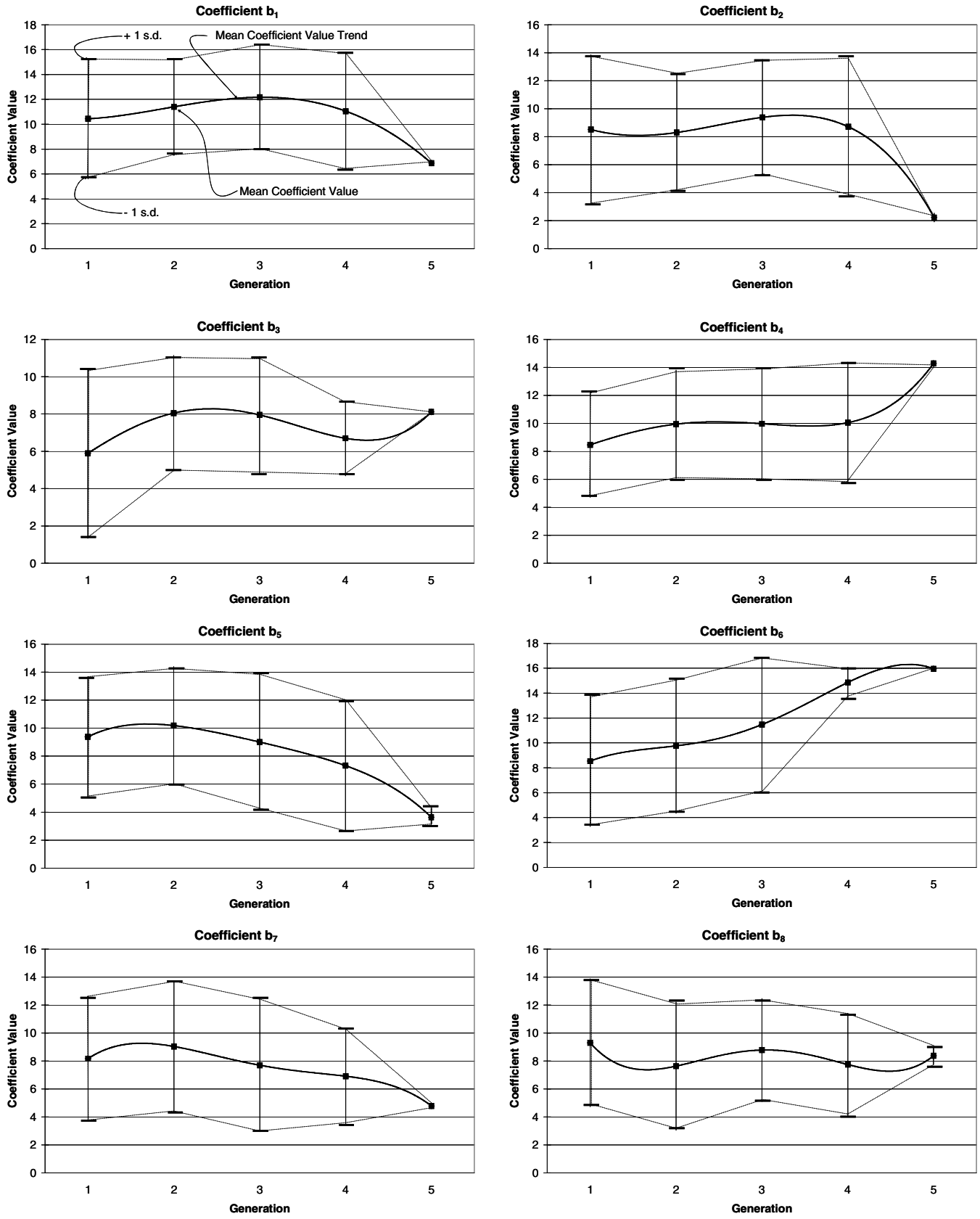


Fig. 6. Convergence of utility coefficients for household 1.

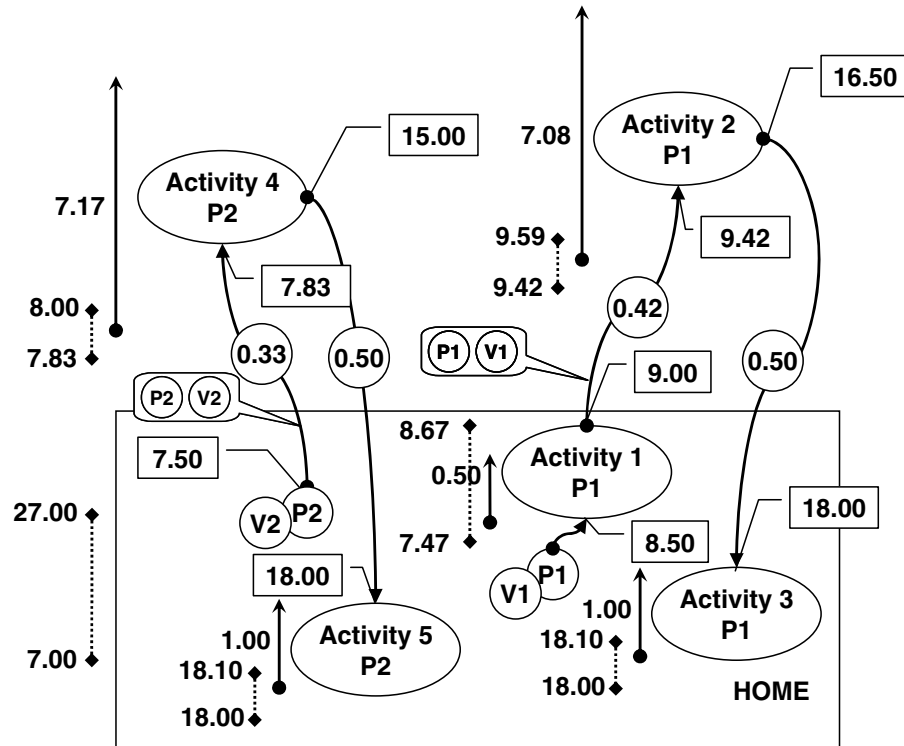


Fig. 7. Observed activity pattern for household 2.

of constraints is further evident in the trace of the convergence of utility coefficients shown in Figure 12, which shows the maximum and minimum values for each coefficient estimate at each generation, together

with its mean across the population of 20 (note that in these results, without loss in generality, the coefficient values have been scaled [0, 1]). We note in particular that the relatively sharp convergence shown in

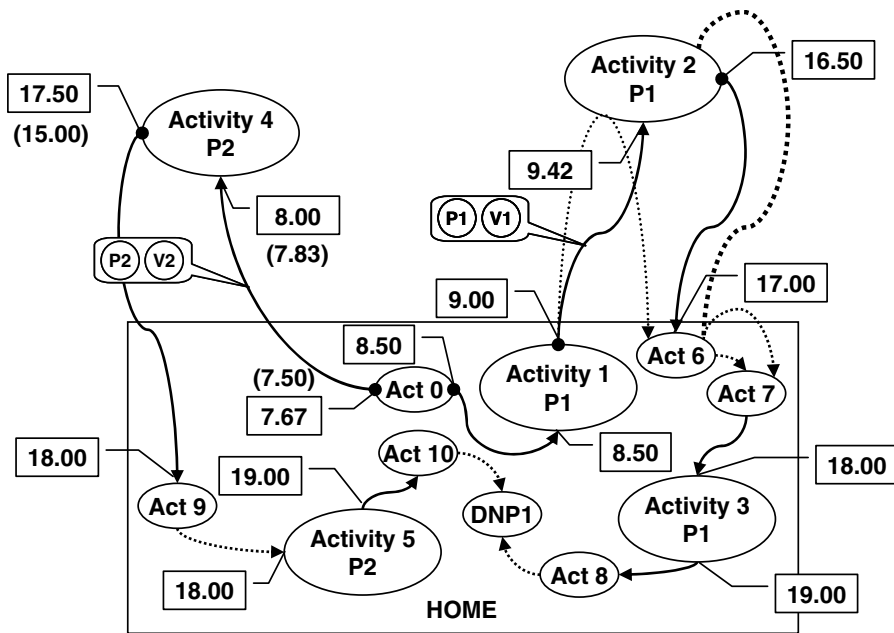


Fig. 8. Best HAPP-predicted pattern for household 2, after generation 5.

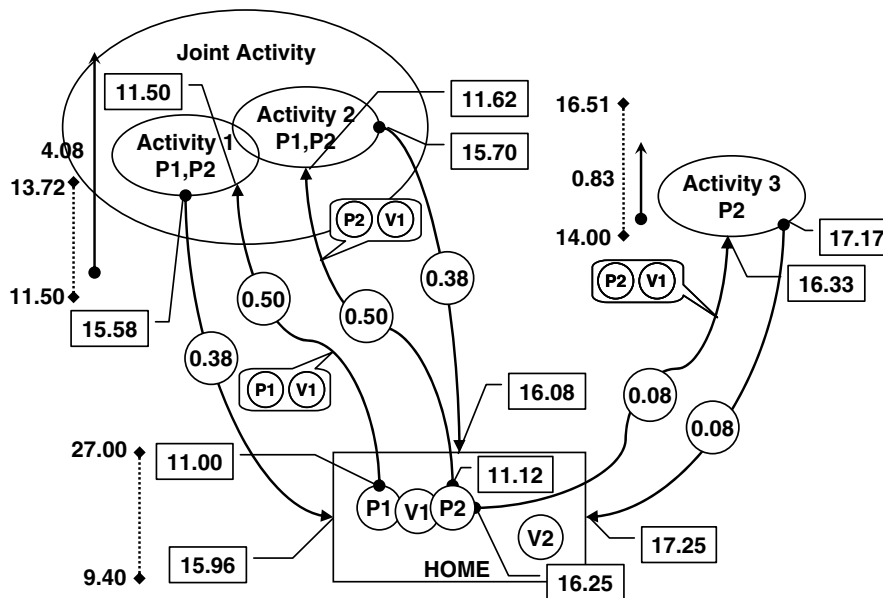


Fig. 9. Observed activity pattern for household 3.

the previous examples of single households in which the utility weights are free to vary across individual households, that is, expressed by $\hat{\mathbf{B}}_i$, in most cases is not apparent under the homogeneity restriction. The implication here is that an assumption more akin to the heterogeneity assumption in the coefficients that underpins mixed logit may be a more attractive option. Moreover, the results of the estimation open questions other than the performance of the GA; it has been noted

that convergence in not particularly impressive, indicating perhaps that the constraints (rather than utility considerations) dominate the choice process—an important result in itself in that it may signify that estimation appropriate for such models may more properly be focused on estimating/identifying the extent of constraints. For example, the apparent poor convergence of coefficients b_2 and b_3 evident in Figure 12 may likely be traced to the fact that the utility components associated

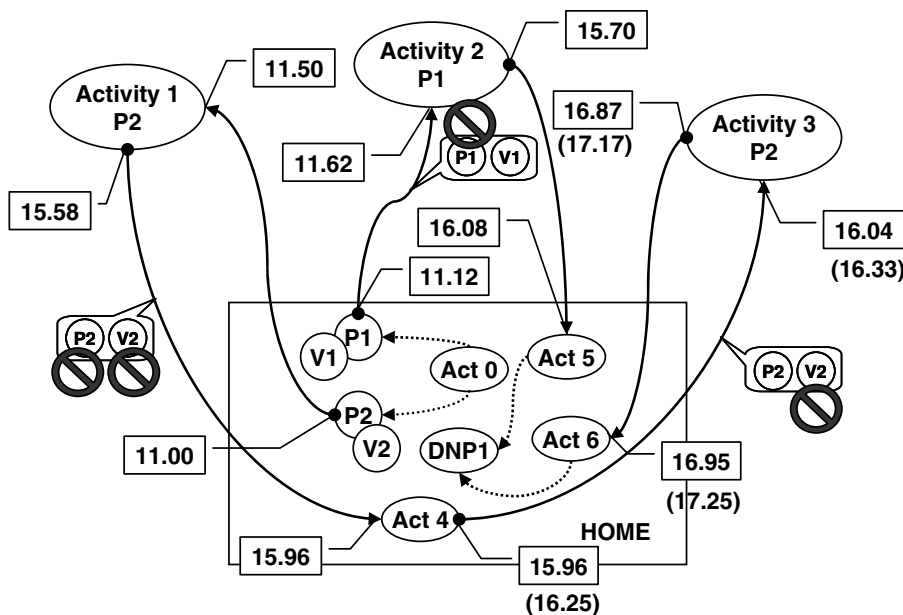


Fig. 10. Best HAPP-predicted pattern for household 3, after generation 5.

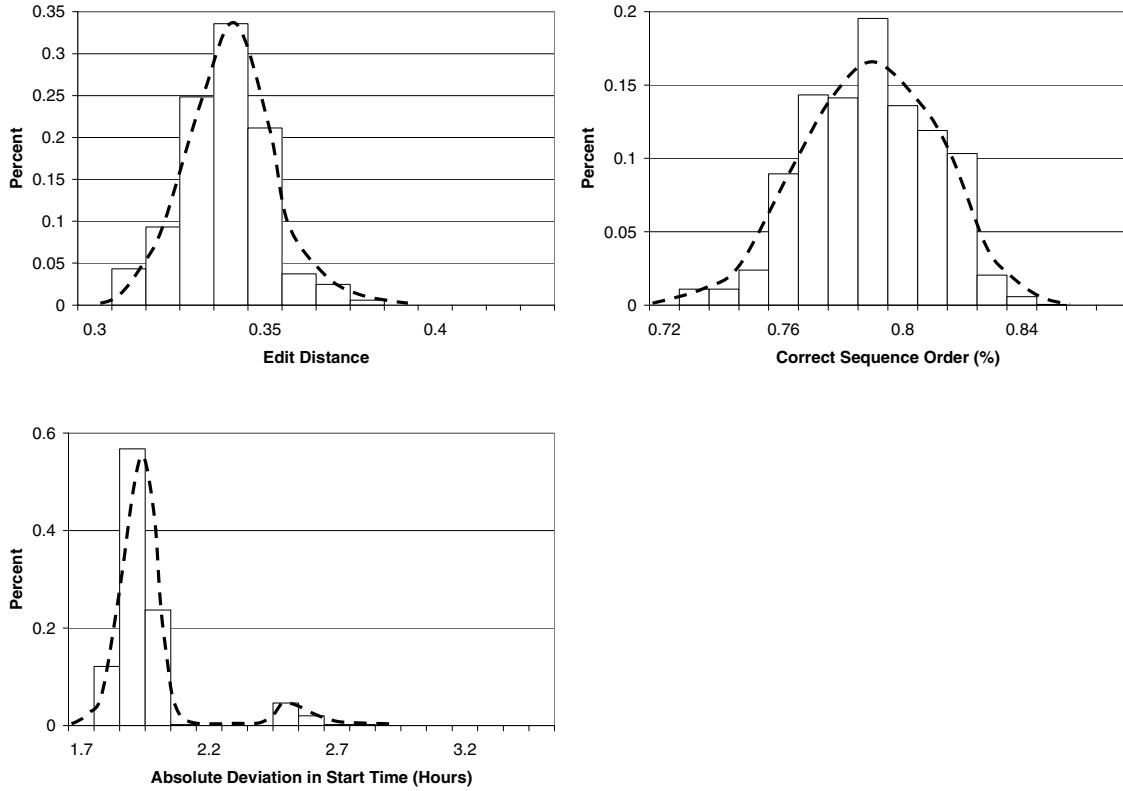


Fig. 11. Histogram of statistical results obtained from the GA.

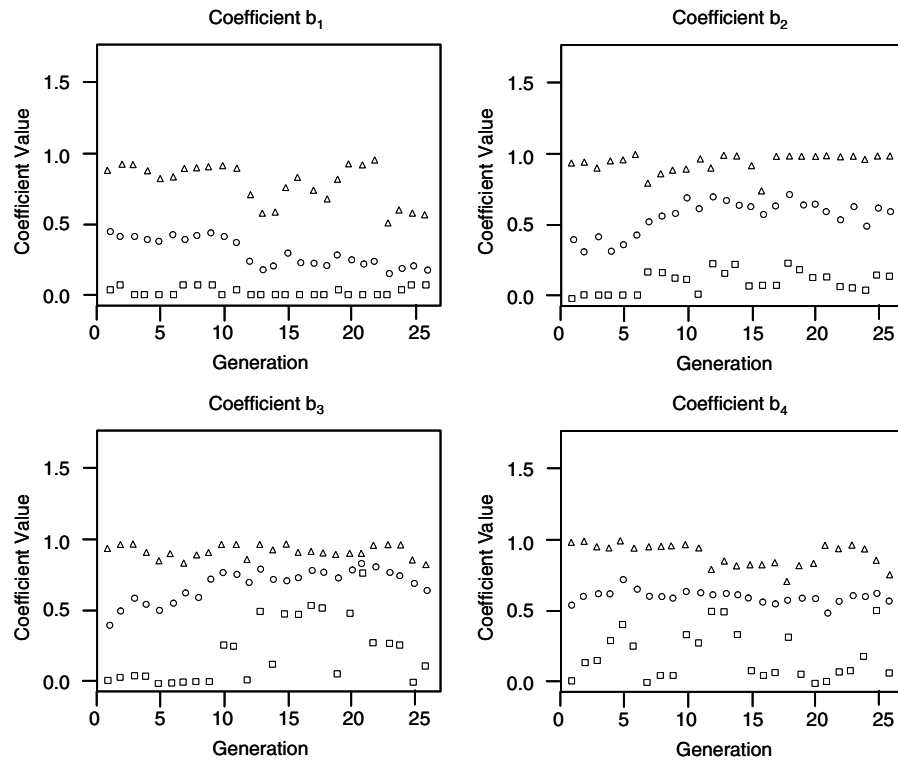


Fig. 12. Convergence properties of utility coefficients via the GA.

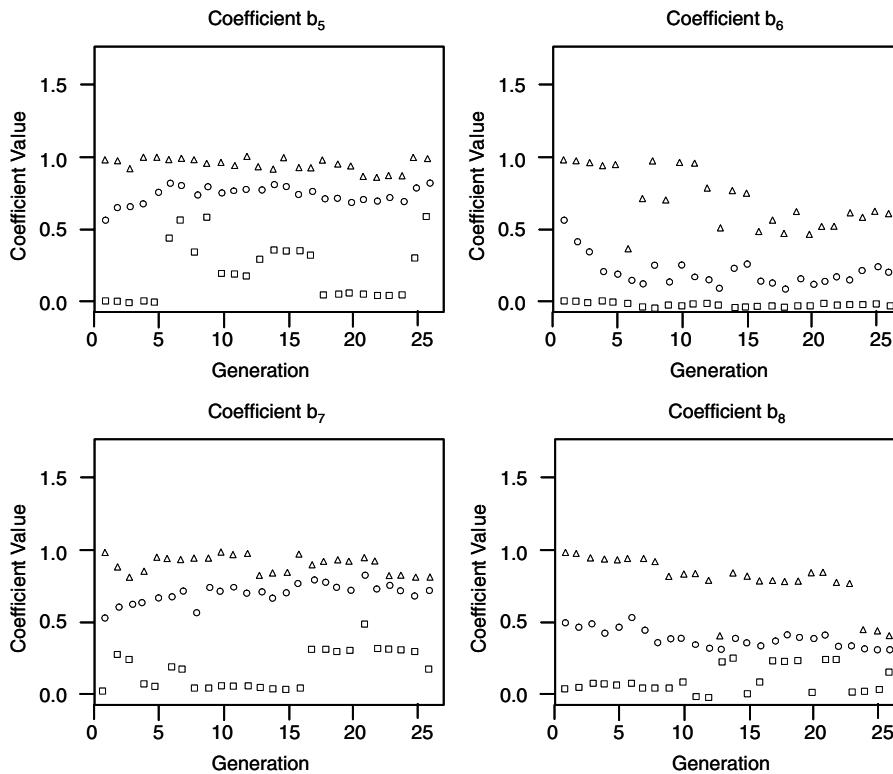


Fig. 12. Continued

with coefficients b_2 , b_3 , b_6 , and b_7 have little (or no) impact on the decision process; unfortunately, without well-defined parametric statistics, standard tests of the coefficients are unachievable.

The overall best fitness score among the candidate $\hat{\mathbf{B}}$'s at any particular generation, followed a pattern similar to that displayed for the individual households of the previous section—most of whatever improvement achieved was captured within the first five generations. This pattern is exemplified by the results for “edit distance” shown in Figure 13; here, improvement in the minimum value was obtained during the first five generations, followed by no further improvement until the 24th generation.

After these 25 generations, the “best” estimates for the components of $\hat{\mathbf{B}}$ were determined as:

$$\hat{\mathbf{B}}' = [0.210 \ 0.221 \ 0.770 \ 0.619 \ 0.875 \ 0.393 \ 0.839 \ 0.306] \tag{8}$$

The components of $\hat{\mathbf{B}}'$ represent the estimates of the relative importance of each of the eight components, that is, the components of $\mathbf{Y}(\mathbf{X}_i)$, comprising the total disutility $\mathbf{Z}(\mathbf{X}_i)$ of an activity/travel pattern to the households. Applying these common utility weights to

the utility functions of the 65 households (130 individuals) in the sample produced the distribution of the three “goodness-of-fit” measures across the sample shown in Figure 14.

We judge these errors to be surprisingly good, considering the complexity of the problem being addressed;

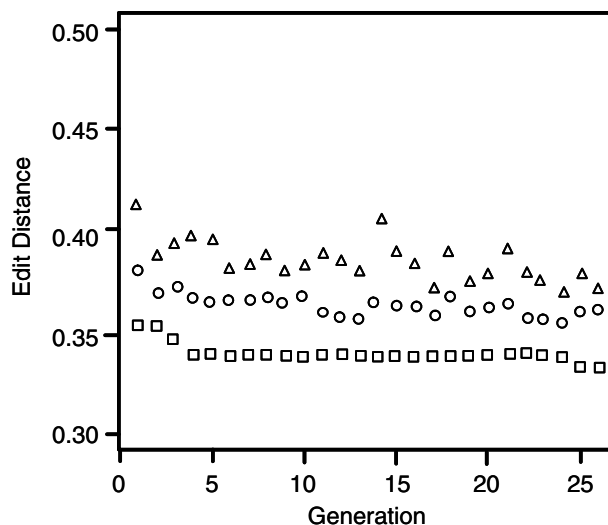


Fig. 13. Convergence of fitness score for “edit distance” to its minimum value.

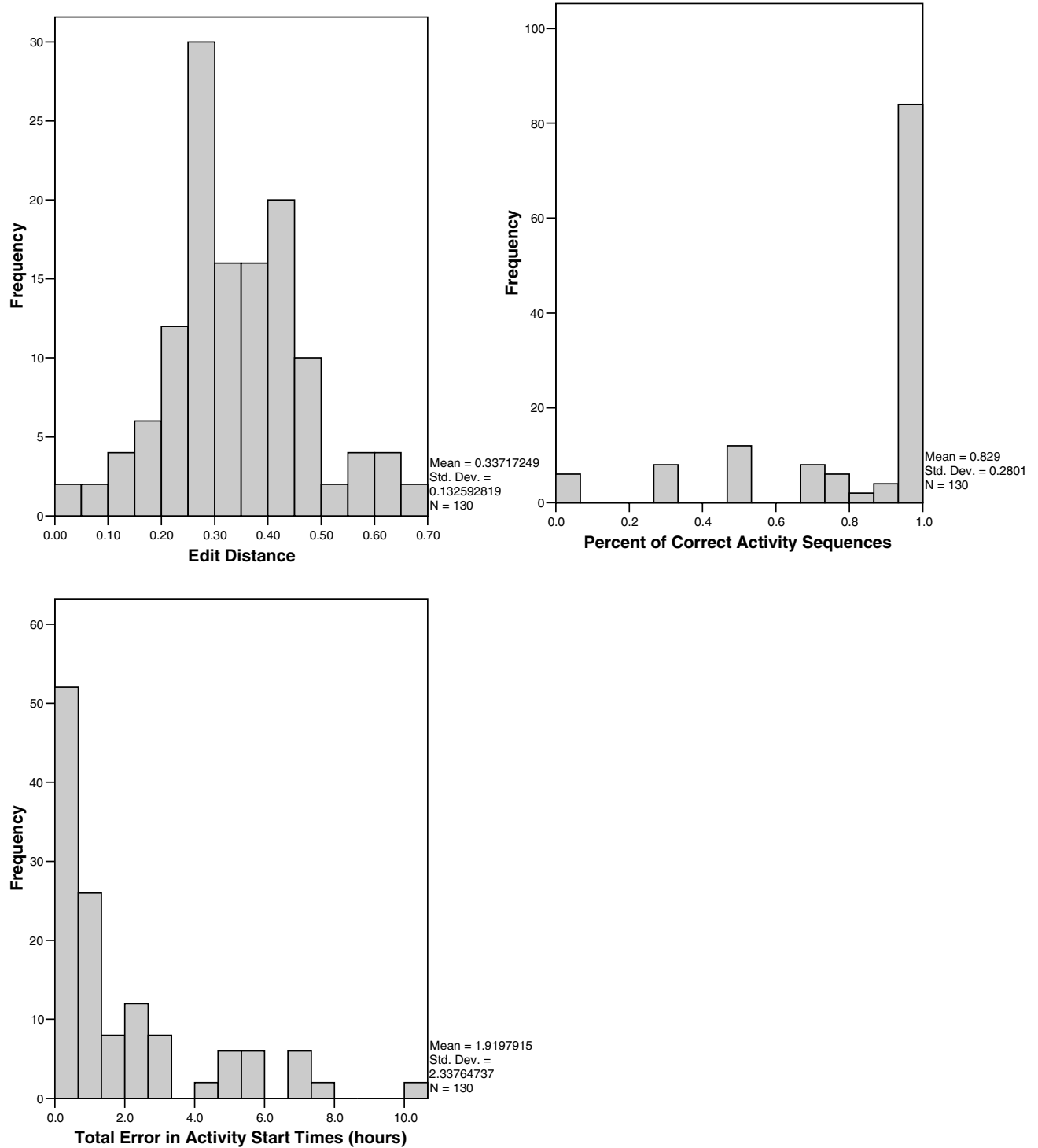


Fig. 14. Distribution of errors in estimation across sample.

that is, the prediction of the combined (1) assignment of activities to members within the household, (2) the start times of those activities, (3) the order of performance of the activities, and (4) the travel linking those activities (e.g., number of tours, particular sojourns compris-

ing tours, etc.). The mean edit distance for the sample was about 0.3; the vast majority of activity sequences were predicted correctly; well over 50% of the sample had little or no error in the predicted start times of their various activities.

5 CONCLUDING REMARKS

Despite the inherent overwhelming complexities of treating the “whole” of travel, the so-called “activity-based” approach has begun to be embraced by the mainstream of transportation researchers, and also practitioners, as offering a viable practical paradigm for travel demand modeling. Already, a number of operational models have been deployed and tested; see Guo and Bhat (2001) for an excellent review and contrasting of most of these models. Some have actually been integrated into the travel demand forecasting practices of metropolitan transportation agencies. For example, the Florida Activity Mobility Simulator (FAMOS) “microsimulation” model system (Pendyala, 2004) has been estimated and calibrated using the 2000 Southeast Florida Household Travel Survey data set. And, the New York Best Practice Model (NYBPM) has been used on a wide range of regional studies to simulate travel patterns (Chiao et al., 2006). In Texas, the comprehensive econometric microsimulator for daily activity-travel patterns (CEMDAP), is currently being evaluated in an application to the Dallas-Fort worth Metropolitan area. Vovsha et al. (2004) provide a comprehensive summary of the progress in moving activity-based modeling approaches to practice.

Although we believe that the procedures reported here represent an important step in moving activity-based travel approaches beyond descriptive analysis toward a foundation for practical demand modeling, there nonetheless remain significant challenges. In the current implementation we have addressed the inference of utility coefficients as a “common” set of coefficients that can be applied to an entire sample of activity patterns with the result that the predicted activity patterns are closest to the actual activity patterns, as determined by a set of three distance measures. This was accomplished by defining a fitness score as some weighted average of “edit distance” and “temporal alignment” across all sample points. Because the total number of possible multidimensional operation sets is equal to the product of the numbers of all optimal paths in the unidimensional sequence alignment method for all attributes, the multidimensional sequence alignment method may be very time-consuming when the number of activities considered is larger than those considered in the test cases reported here; assuming that the operation cost for one insertion and one deletion is the same as one substitution in the unidimensional sequence alignment method, the number of all optimal paths for a pair of sequences with length m will have the complexity of $o(m^m)$. Additionally, the time sequence comparison technique that has been employed, which simply offsets activities along the time dimension to maximize the overlapped time or number of overlapped activities, is quite simplistic. For

purposes of illustration, we have used a fitness score that is simply the difference between the “sequence distance” (i.e., the number of steps to equate two sequences) and the degree of overlap along the temporal axis. Though the fitness score is a scaled value, it has no practical meaning. Moreover, there are no well-defined statistical properties for any of the measures that would allow statistical confidence estimates for the results.

Finally, despite its complete description (and analytical management) of the “hard” physical constraints imposed on a household’s collection of activity/travel decisions, the HAPP model has no explicit representation of such “soft” constraints as each member’s role in the household, or their particular predilections and idiosyncrasies that may be imposed on top of these physical constraints. One promising avenue for incorporating these considerations may be in adapting some of the procedures identified with the relatively new field of uncertain programming (Liu, 1999).

ACKNOWLEDGMENTS

This research was supported, in part, by a grant from the USDOT University Transportation Center of the University of California, with matching funds provided by the Division of Research and Innovation of the California Department of Transportation. Their financial support is gratefully acknowledged.

NOTE

1. For example, greater relative weight may be placed on correctly replicating the patterns of large households with many activities, or on those that reveal work trips taken during peak hours, or on those for which activity diary data are judged to be more reliable than others.

REFERENCES

- Arentze, T. A., Hofman, F., Joh, C.-H. & Timmermans, H. (2002), Activity pattern similarity: a multidimensional sequence alignment method, *Transportation Research Part B: Methodological*, **36**, 385–403.
- Bhat, C. R., Guo, J. Y., Srinivasan, S. & Sivakumar, A. (2004), A comprehensive econometric micro-simulator for daily activity-travel patterns, *Transportation Research Record*, **1894**, 57–66. Q3
- Cambridge Systematics. (1996), *Data Collection in the Portland, Oregon Metropolitan Area, Travel Model Improvement Program*. Report prepared for the U.S. Department of Transportation, Oakland, CA.
- Chiao, K. A., Bhowmick, S. & Mohseni, A. (2006), Lessons learned from the implementation of NY activity-based travel models, *Presented at the Innovations in Travel Modeling Conference*. Q4
- Guo, J. Y. & Bhat, C. R. (2001), *Representation and Analysis Plan and Data Needs Analysis for the Activity-Travel*

- 1
2
- 3 System, Report No. 0-4080-1, Center for Transportation
4 Research, The University of Texas at Austin, TX.
- 5 Holland, J. H. (1992a), *Adaptation in Natural and Artificial*
6 *Systems*, 2nd edn., MIT Press, Cambridge, MA.
- 7 Holland, J. H. (1992b), Genetic algorithms, *Scientific American*,
8 July, 66–72.
- 9 Liu, B. (1999), *Uncertain Programming*, John Wiley & Sons,
10 New York.
- 11 Michalewicz, Z. (1992), *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, New York.
- 12 NuStats. (1995), *Final Report on Oregon and Southwest Washington Household Activity and Travel Survey*, Austin, TX.
- 13 Pendyala, R. M. (2004), *Phased Implementation of a Multimodal Activity-Based Travel Demand Modeling System in Florida, Volume I: Technical Documentation*, Prepared for the Florida Department of Transportation, Tallahassee, FL.
- 14 Recker, W. W. (1995), The household activity pattern problem: general formulation and solution, *Transportation Research*, **29B**(1), 61–77.
- 15 Recker, W. W. (2001), A bridge between travel demand modeling and activity-based travel analysis, *Transportation Research, Part B: Methodological*, **35**, 481–506.
- 16 Recker, W. W. & Parimi, A. (1999), Development of a microscopic activity-based framework for analyzing the potential impact of TCMS on vehicle emissions, *Transportation Research, Part D: Transport and Environment*, **4**, 357–78.
- 17 Recker, W. W., Chen, C. & McNally, M. G. (2001), Measuring the impact of efficient household travel decisions on potential travel time savings and accessibility gains, *Transportation Research, Part A: Policy and Practice*, **35**, 339–69.
- 18 Rudolph, G. (1994), Convergence properties of canonical genetic algorithms, *IEEE Transactions on Neural Networks*, **5**(1), 96–101.
- 19 Vovsha, P., Bradley, M. A. & Bowman, J. L. (2004), Activity-based travel forecasting models in the United States: progress since 1995 and prospects for the future, in *Paper presentation at the EIRASS Conference on Progress in Activity-Based Analysis*, Vaeshartelt Castle, Maastricht, Netherlands.
- 20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56

QUERIES

- Q1** Author: Michalewicz, 1994 is not mentioned in the list. Please check.
- Q2** Author: Please check, Equation 6 is given twice. Do we need to renumber the equations?
- Q3** Author: Please provide citation for reference Bhat et al., 2004.
- Q4** Author: Please provide the location in reference Chiao et al., 2006.
- Q5** Author: Please provide citation for reference Michalewicz,1992.
- Q6** Author: Please provide citation for reference Recker, 2001.