MODELING TRAVEL UNDER CONSTRAINED CHOICES

by

Yacov Zahavi
Mobility Systems, Inc.

Distributed at
The Conference on Analysis Methods for the 1980s
Transportation Research Board
October 3-7, 1982
Easton, Maryland
INTRODUCTION

Among the purposes of this conference is the search for new approaches to travel modeling, in order to answer significantly different planning requirements in the 1980s from those of the recent past.

A prerequisite for the development of new approaches is the identification of the problems that prevent the 1970s models from being fully useful in the 1980s. Hence, the first part of this paper identifies several major conceptual problems associated with the conventional travel models.

The second part presents the basic structure of a new travel model that answers to the conceptual problems of conventional models. The third part of the paper details some of the empirical data that support the new approach. The paper concludes with several recommendations for further tests of the new model.

SOME CONCEPTUAL DIFFICULTIES

Some of the major conceptual difficulties associated with conventional travel models are detailed below. The purpose of the following list is to emphasize the prevailing general approach to modeling travel, rather than to single out a specific model. While doing so, it became increasingly evident that conventional urban travel models display many conceptual difficulties, most of which seem to be in conflict with basic principles in system theory (Kalman, 1973).

"Conventional models" refer in this paper to all operational travel models that display one or more of the following conceptual difficulties:

- Transferability

  The essential purpose of travel models is prediction. Therefore, a model based on observations in one city must be shown to have the ability to 'predict' travel conditions in other cities at the same point of time if it is to be regarded as a model able to predict travel in the same city at a different point of time.

This, however, is not the case with conventional travel models; these models are found to be non-transferable when the cities are different in size (e.g., Atherton, 1976; Caldwell, 1980), nor even transferable between different parts of the same city (Galbraith, 1982). Part of the problem in this case is that the coefficients of a model, fitted to cross-sectional data, are assumed to remain fixed over time for each city while, at the same time, such a model usually must be re-calibrated in each separate city. However, the coefficients cannot be non-transferable (over space) but transferable (over time) at the same time. Indeed, coefficients were found to change over time also in urban structure models (Wilson, 1981).
In short, as long as a travel model is not fully transferable over space, between cities, it should not be regarded as transferable over time in any one city.

- **Calibration and Validation**
  Conventional travel models are calibrated to the observed travel choices. Thus, both the independent and dependent variables must be known before such models can be calibrated. In most cases such a model is then validated by its statistics and by its ability to reproduce the same observations to which it was fitted. This may be regarded as a tautological process. Furthermore, even a perfect statistical fit to cross-sectional data does not ensure temporal transferability of the model, namely its ability to predict correctly.

  In short, the conventional calibration process of conventional travel models, which is a lengthy and costly process, does not necessarily prove, nor validate, the model; ten different models, including contradictory ones, can be calibrated to the same data set and still pass statistical and sensitivity criteria.

- **Causality**
  Causality in conventional travel models is typically assumed *a priori*. For example, it is typically assumed that car availability per household increases trip generation. Namely, car ownership is the cause, while more trips is the effect. However, it might be also argued legitimately that the need for more travel generates car ownership levels.

  The point is that within a complex travel system principal travel components can be both cause and effect, depending on the feedback step.

- **Stable Parameters**
  Conventional models are based on trips. However, the definition of a trip, and its related trip distance, trip time and trip cost, is ambiguous to a large extent since trips are linked/chained/clustered and combined into "toours" in various ways during the calibration phase of the models. Thus, the basic travel data in a city can vary according to the chosen definition of a trip, resulting in different models for the same city. This problem may contribute to the non-transferability of a model between different cities.

  In short, a model should be based on components that remain unchanged by any definition.

- **Organization Process**
  A conventional travel model is divided into many sub-models; car ownership sub-model, trip-generation sub-model, trip mode-choice sub-model, and trip-distribution sub-model. Each sub-model is further sub-divided by trip purpose and/or mode, with no explicit interactions between the various sub-models. Moreover, each sub-model is based on a large number of independent variables (e.g., 21 different variables in one dis-
aggregate car-ownership sub-model), which are used in various combinations for different population segments.

An additional problem is that many of the independent variables are used repeatedly and/or alternately in separate sub-models. For example, income is used as one independent variable of several in the car ownership sub-model, while later car ownership and income are used as two of the independent variables in the trip-generation sub-model, and so on.

The point is that conventional models are open-ended, composed of many separate, non-interacting, parts. What is required is one organizing principle, encompassing all travel components simultaneously within one interacting system.

- **Equilibrium vs. Disequilibrium**

  Conventional travel models are usually based on the assumption that travel demand is in equilibrium with system supply. Thus, by definition, each alternative scenario must reach, or at least approach, equilibrium between demand and supply.

  Such assumptions are especially troublesome in the so-called quasi-dynamic models, where equilibrium is assumed at the end of each time-increment, say 3 years.

  However, it is equally valid to say that it is the amount of possible disequilibrium associated with alternative futures which generates forces that dynamically change urban structure, often in unexpected ways. Thus, the travel system might be in disequilibrium conditions at the end of each assumed time increment, namely in contradiction with the starting assumptions.

  In short, a model should be able to identify and quantify possible disequilibrium conditions, a requirement which conventional travel models do not satisfy.

As can be seen from the above examples, conventional models, even joint-probability ones, display one or more of the conceptual difficulties, difficulties which put a question mark on the usefulness of such models for predicting travel.

A different approach to the modeling of travel is presented in the next section, an approach that solves, or rather bypasses, all the above conceptual difficulties associated with conventional travel models.

**THE UNOT APPROACH**

Most of our choices are arrived at under some constraints, such as physical, emotional, religious, or economic. A useful way of structuring a travel model under constraints is by optimizing an objective function under explicit constraints, such as those of travel time and money. Indeed, this approach serves as the basis for two classes of models, the utility-maxim-
zation and the entropy-maximization models. However, in both cases several conceptual problems prevent the models from being fully useful. First, they are based on trips and thus are affected by the definition of a trip. Second, the maximization process is used only at the phase of structuring the mathematical format of the model, while the model itself has to be calibrated to observed choices the same way as any other conventional model. Consequently, the utility and entropy maximization models still suffer from most of the conceptual difficulties mentioned above.

The UMOT, or Unified Mechanism of Travel, is an urban travel model based on a multi-loop feedback approach in which all travel components, transport system, and urban structure interact simultaneously (Zahavi, 1979, 1981, 1982). The model is based on just one hypothesis, namely that travelers attempt to maximize the benefits (utility) of their accessibility to spatial and economic opportunities under explicit constraints of travel time and money budgets.

Perhaps the best way to explain the UMOT approach is by describing how it bypasses all the conceptual pitfalls that affect conventional models:

- **Transferability**

  The allocation of daily travel time per traveler and daily travel money per household in urban areas were found to display consistent regularities, transferable over both space and time not only within but also between countries. Thus, the constraints used in the UMOT model, as well as all other basic relationships (e.g., travelers per household), are only those that were found to be transferable over space and time. Consequently, the model itself is fully transferable.

- **Calibration and Validation**

  In the UMOT model no desired output is ever calibrated to the observed values. The outputs are the predicted choices, which are then compared with the observed choices - not fitted to them - for the model's validation.

  For example, the process can be started by assuming that each and every household in the urban area owns, say, 5 cars. Such an assumption, of course, is absurd. Nonetheless, the travel system converges rapidly to the observed car ownership levels, and all other travel choices, by household segments.

- **Causality**

  In the UMOT model there are no assumptions about unilateral, fixed, causality. The process is based on a systemwise approach, where all travel components interact with each other and with the transport system through a simultaneous dynamic feedback process. Thus, each component can be both cause and effect, depending on the feedback step.

- **Stable Parameters**

  The UMOT model is based on travel components that remain unchanged by any definition, which are the total daily travel components, such as the daily travel distance, and the daily travel time and money expenditures, per traveler/household.
Organization Process

In the UMOT model there is only one organizing principle which operates the travel system: an objective function which represents the individual travelers/households attempts to maximize their benefits of accessibility to spatial and economic opportunities under their travel constraints.

In contrast to conventional optimization models, the maximization process in the UMOT model is activated during each iteration of the model, encompassing all principal travel components simultaneously.

Equilibrium vs. Disequilibrium

One of the outputs of the UMOT model is the amount of potential disequilibrium affecting different population segments. The disequilibrium is measured by the amounts of daily travel time and/or money that travelers have to spend above the minimum (asymptotic) values. Put another way, travelers that are forced to spend more than the minimum values of time and money will attempt to change their travel/location patterns in the urban area, under their constraints, such as by changing destinations in the short run, changing modes and car ownership levels in the medium run, and changing residence/job locations in the long run.

As can be seen from the above examples, there are some significant differences between the UMOT travel model and conventional travel models. Such differences are reflected not only in the concepts underlying the models, but also in the way the models are activated.

Perhaps the best way of describing the systemwise approach used in the UMOT model is by detailing the principal feedback steps of the process. The process can be started, as already mentioned above, by assuming that each and every household in the urban area owns 5 cars. Such an assumption, of course, is absurd, but at this stage the travel constraints take over and drive the travel system, based on given unit costs, to result in (i) the estimated travel distance by each mode, and (ii) the estimated car ownership levels. (iii) The interaction between the estimated number of cars and a given road network results in new unit costs of travel, which are fed back into the travel distance phase, and (iv) repeating the process by iterations results in the rapid convergence of the travel system (where all travel components, including the travel budgets, interact with each other) to outputs which agree with the observed travel and car ownership levels in the urban area.

Consequently, there is no need to calibrate separate sub-models for the travel components, such as for car ownership levels; the process responds to any input, even an absurd one, and it adjusts the system to converge to expected values within the organization principles. Borrowing an expression, the UMOT process can be termed as a "self organizing system". In all cases tested until now, the expected values of travel components were found to match the observed ones.

Table 1 presents a simplified flow chart of the UMOT travel model, showing the principal feedback processes. Further developments are the extension of the UMOT travel model to include urban structure, as well as intercity travel (Zahavi, 1982).
Table 1: Flow Chart of the Interactions between Travel Demand, System Supply and Car Ownership, the UMUT Model

- Input/Output flow
- Interaction. Effect of Input on Output is expressed by ⊗ or ⊘
- → Feedback

The next section presents empirical data that support the basic concepts behind the UMUT model.
EMPIRICAL DATA

Figures 1 and 2 show the daily travel time per traveler frequency distributions, at disaggregate levels, in Baltimore 1977, stratified by six income groups (Zahavi, 1982).

FIG. 1 TRAVEL TIME/TRAVELER DISTRIBUTIONS

The slight differences between the distributions in Figure 2 is accentuated in Figure 3, where the travelers are stratified by auto ownership levels.

FIG. 2 TRAVEL TIME/TRAVELER DISTRIBUTIONS

FIG. 3 TRAVEL TIME/TRAVELER DISTRIBUTIONS
The full implications of such characteristics can be appreciated when adding Figure 4, showing the daily travel distance distributions by auto ownership levels; the higher the auto ownership level (and speed), the less travel time is spent for more travel distance.

The effect of speed on the daily travel distance per average traveler is shown in Figure 5, where the same relationship holds whether the travelers are stratified by auto ownership, income, or size.

Such relationships can explain the reluctance of travelers to transfer from auto to bus travel; they will then have to spend more time for less travel distance. Such relationships also suggest that a transfer of trips between modes having different speeds are not a one-to-one transfer, since trips may be gained, lost, or their characteristics changed by such transfers.

Figure 6 suggests that the above regularities are also transferable spatially, among four different cities, two in the U.S. and two in the U.K., where the slight shift of the distributions towards the right (i.e., longer travel times) is attributed to the differences in the door-to-door speeds in the four cities (Zahavi, 1982a). Similar consistent and transferable regularities were also found for the daily travel money expenditures as a proportion of income.
The same consistent and transferable regularities were also displayed over time. For instance, the daily travel times per average car traveler in Washington, D.C. were 1.09 and 1.11 hours in 1955 and 1968, respectively, although door-to-door speeds increased by 24 percent during this period. In Twin-Cities the values were 1.14 and 1.13 hours in 1955 and 1970, respectively, although door-to-door speeds increased by 33 (1) percent (Zahavi, 1979a). Thus, the saved travel times were not saved, as such, but were traded off for more travel distance. Put another way, a model should not stop at the point where saved travel time is estimated, but it should also consider the feedback effects of such savings.

The same consistent regularities were also displayed at the nationwide level: the daily travel time per average traveler in the U.S. in 1970, derived from the 1970 Nationwide Personal Transportation Study, was 1.08 hrs. (Zahavi, 1974). The value in 1977 was found to be 1.02 hrs., virtually unchanged over a period of 7 years which had seen significant changes in oil prices, car ownership and travel characteristics.

In summary, the transferable regularities shown above go beyond any coincidence, and they lend strong support to the UMOT approach.

RECOMMENDATIONS FOR FURTHER RESEARCH

The UMOT process has been tested using a number of different utility functions. The results of these tests show that the utility weights derived from the Baltimore data produce good results when applied to the London data. Conversely, the utility weights derived from the London data produce equally good results when applied to the Baltimore data (Zahavi, 1962). Thus, if one regards the development of the utility weights as a "calibration", then, in each case, the calibration in one city was valid for the other. This kind of tests should be extended to other cities to further test the transferability of the models.

A second outcome of the tests and data analysis was that there is more variability in the observed relationships at the lower end of the income scale than at the middle or upper end, and this has some effect on the aggregation used in the model. Preliminary analyses have been made of the joint distributions of income, travel distance, travel speeds, house-
hold size, etc., which indicate wide dispersions about the observed regularities in the lowest income class, but becoming very stable at all other income classes.

A concentrated effort should be made, starting with the Baltimore data, to estimate the joint distributions (at least in significant pairs) and to determine their implications for the stratifications to be used in UMDT, as well as for the utility functions which must, when optimized, produce the observed travel regularities.

* * *

(Note: This is a summary paper; some parts of it already appeared in previous reports/papers.)

REFERENCES


