Paper No. 02-3668

# Microsimulation in Travel Demand Modeling

## Lessons Learned from the New York Best Practice Model

Peter Vovsha, Eric Petersen, and Robert Donnelly

**Microsimulation is increasingly assuming a major role in the advancement of demand-modeling practice. At the same time, it is attracting growing attention from the larger transportation-planning community. Four basic advantages of microsimulation versus conventional fractional-probability models are examined. The first is the technical advantage related to computational savings in the calculation and storage of large multidimensional probability arrays. The second is the meaningful advantage gained in the explicit modeling of various decision-making chains and time–space constraints on individual travel that allows for behavioral realism in the demand-modeling procedure. The third relates to the variability of microsimulation outcomes, which can yield full information about the distributions of the travel demand statistics of interest rather than single deterministic estimates or average values. As soon as constraints are introduced into the modeling framework (which often is done at the destination choice stage), competition arises, although generally it has been ignored in standard models. Microsimulation has the potential to handle this competition over work attractions and other travel activities in a meaningful fashion, which is the fourth advantage. These four advantages of microsimulation are discussed in light of the recent development and application of the New York best practice model, a microsimulation demand-modeling system for the New York–New Jersey–Connecticut metropolitan area.**

Microsimulation should be viewed as a natural extension of the disaggregate modeling technique, rather than an alternative approach to the conventional demand models, as it is sometimes presented. Theoretically, the microsimulation concept is based on the individual (or household) unit of behavior—just as in any disaggregate model. However, the microsimulation process generates discrete choices of the individual (trip purpose, destination, mode, time of day) rather than an array of probabilities for a population segment associated with each available alternative. The ability to simulate travelers individually allows for consideration of complex linkages across multiple trips, ultimately resulting in a better estimation of real-world travel behavior.

Microsimulation requires that a core probabilistic model be developed and calibrated. In general, there is not much difference between the statistical estimation of a standard disaggregate model and that of a model to be applied in the microsimulation context. However, during application, a Monte Carlo approach is applied to simulate discrete choices at the individual level, rather than relying on the aggregation of fractional probabilities along travel demand dimensions. In most applications these individual realizations from microsimulation are aggregated before network loading to obtain traffic and transit statistics, though the goal of some research is the ability to assign these discrete trips directly on a network.

PB Consult, Parsons Brinckerhoff, 5 Penn Plaza, 17th Floor, New York, NY 10001.

The most logical basis for the Monte Carlo technique is a fully disaggregate model built and calibrated on the individual (or household) level. Such a model application would be based on the synthesized population, that is, a full list of household and related people in each zone with attributes relevant for the calculation of choice probabilities (household size, income group, automobile ownership, person category). This way it would fully account for variation of the household composition in the multidimensional space of attributes.

This article is organized as follows: first, the four basic advantages of microsimulation are formulated, and then each advantage is analyzed in detail in its own section. The examples are drawn from the recent application of the New York–New Jersey–Connecticut metropolitan model developed for the New York Metropolitan Transportation Council—referred to as the New York best practice model (NYBPM). The term best practice model reflects the practical orientation of the project and is in reference to previous regional models. Finally, concluding remarks and recommendations for future research are outlined.

## ADVANTAGES OF MICROSIMULATION

It should be stressed here that the application of the Monte Carlo technique on top of the core probabilistic model itself does not automatically improve either the behavioral structure or the accuracy of the transportation model. If the core probabilistic model can be applied simultaneously across all travel demand dimensions, a Monte Carlo game on top of it simply follows the basic travel statistics already available from the core model.

Microsimulation does, however, have several major advantages over standard fractional-probability models. First, microsimulation offers substantial savings in the calculation and storage of multidimensional fractional-probability arrays during model application.

Second, microsimulation allows for the explicit formulation of various chained decisions and time–space constraints on individual travel behavior. To truly take full advantage of microsimulation, the core probabilistic model must be restructured to take chained travel behavior into account. The estimation stage will require more effort because of the complication of the core probabilistic model (conditioning and truncation).

The third potential advantage of microsimulation relates to the explicit modeling variability of travel demand, rather than of average values. The last property can be exploited as a significant advantage of microsimulation in certain circumstances, whereas it may be viewed as a problem in other cases.

As soon as constraints are introduced into the modeling framework, often done at the destination choice stage, competition arises, though it has generally been ignored in standard models. Microsimulation has the potential to handle this competition over work attractions and other travel activities in a meaningful fashion, which is its fourth advantage.

## MULTIDIMENSIONAL PROBABILITY ARRAYS

Microsimulation offers substantial savings in the calculation and storage of multidimensional fractional-probability arrays during model application. That feature is especially relevant for large metropolitan areas characterized by thousands of traffic zones, dozens of travel modes and mode combinations, and a nonhomogeneous population, in which the number of travel-relevant strata may be well over 100 distinct combinations. That was the primary reason microsimulation was incorporated in the NYBPM.

Above a certain size threshold, microsimulation seems the only realistic approach regardless of the mathematical complexity of the core probabilistic model. For large metropolitan areas, the number of individuals, although in the millions, is significantly lower than the number of probabilistic cells in a multidimensional matrix of modeled travel, easily in the billions, where the number of zones is large and a reasonable level of market segmentation is required. Although progress in both computer hardware and software increases the limits for storage and access of large multidimensional arrays, there is still an order-of-magnitude gap between the desired dimensionality and any conceivable computer capabilities in the foreseeable future.

To illustrate that aspect, the following basic dimensions of the NYBPM can be analyzed in both conventional and microsimulation frameworks:

- Nearly 4,000 traffic zones, yielding a full origin and destination (OD) matrix of almost 16 million cells.
- Ten aggregate travel modes (not including nonmotorized travel, which is treated separately):
  - Drive alone,
  - Shared ride-2 (driver plus one passenger),
  - Shared ride-3 (driver plus two passengers),
  - Shared ride-4+ (driver plus 3 or more passengers),
  - Transit (including bus, subway, and ferry) with walk access,
  - Transit with drive access,
  - Commuter rail (with transit feeder lines) with walk access,
  - Commuter rail with drive access,
  - Taxi, and
  - School bus (for journeys to school only).
- More than 100 population slices based on a Cartesian combination of factors such as the following:
  - Three household income groups (low, medium, high),
  - Four household car sufficiency groups (without cars, fewer cars than workers in the household, cars equal to number of workers, more cars than workers), and
  - Number of household members grouped into three personal categories (worker, nonworking adult, child).
- Six journey purposes (work, school, university, household maintenance, discretionary activity, and non-home-based at-work journeys).
- Four time periods (a.m., midday, p.m., night).

To simplify the presentation we will assume that each journey purpose can be modeled separately and a time-of-day distribution of travel can be modeled by means of constant peak factors. Thus, travel purpose and time-of-day dimensions can be moved out of the modeling framework although the total running time will be multiplied by the number of segments.

Even with those assumptions, the simplest multinomial logit formulation for destination- and mode-choice models leads to an infeasible array of fractional probabilities that will contain 16 billion cells, most filled with extremely small fractional numbers. Such an array could never be explicitly calculated and stored, even though all background models, such as the mode- and destination-choice models, could be calibrated. Any further complication of the modeling structure, such as nesting, linkage of journeys of different purposes in trip chaining, intrahousehold interaction, or introduction of the time-of-day dimension, appears to be completely prohibited.

It does not mean that the entire array is used simultaneously at the same modeling stage. In most cases, choices are modeled independently by segments. In particular, additional segmentation by household/person type (implicitly applied in NYBPM for "packeting" journeys and savings in choice probability calculations) can be explicitly applied along with the segmentation by travel purpose and time of day. Although a deep segmentation can make each subarray manageable, massive run time would be required to repeat the calculations. Additionally, there is still the problem of storing the whole array to produce various reports without rerunning the model. A further consideration is that the current modeling tendency creates more and more interlinkages across segments.

Microsimulation opens a constructive way to effectively handle this multidimensional travel demand, especially for large and complex urban models. Total population of the New York–New Jersey–Connecticut metropolitan area is about 20 million. The total daily number of paired journeys is about 23 million, and the biggest purpose—household maintenance—consists of about 8 million journeys. A file with 8 million records can be efficiently stored and handled using a programming language such as C. Preparing the core probabilities for the Monte Carlo events requires the calculation of several large arrays along certain dimensions; yet there is no need to explicitly calculate and store a full multidimensional array of 16 billion cells.

The NYBPM microsimulation procedure passes through the following basic stages (several technical details are omitted to present the basic concepts more clearly):

- Journey-frequency choice. The journey-frequency choice model produces a file containing all 23 million paired journey records organized into six purposes, with an origin-zone (home) anchor for each individual.
- Destination choice. The destination-choice model is applied to each journey purpose separately. The destination choice model attaches a destination zone to each record based on household and individual characteristics, attributes of the known origin zone, and calculated origin–destination impedances (including mode-choice log-sum) to all 4,000 destinations. The major computational saving from microsimulation is achieved by the fact that there is never a need to calculate and explicitly store a full array of destination-choice probabilities for all origin–destination pairs and population strata. Instead, the destination choice utility is effectively broken into a socioeconomic part (dimensioned 4,000 origins $*$ 100 strata) and origin–destination impedance (dimensioned 4,000 origins $*$ 4,000 destinations $*$ 12 strata). A reduced number of strata (12) are relevant for the origin–destination impedance. These two arrays can

be calculated and stored in the computer memory. During the microsimulation, individual records are processed one by one. A subarray of 4,000 full destination-choice utilities can be easily calculated for each of them. Once the Monte Carlo random pick of the destination zone is completed, the subarray is released from the memory and the next record is processed.

- Mode choice. The mode-choice model is applied for each journey after the origin and destination are known. It is based on the calculation of 10 origin–destination mode-utility arrays; each of them has a 4,000 * 4,000 origin–destination dimensionality. Additionally, as in the destination-choice case, there are stratified origin-based and person-based components of mode-choice utilities. The 12 strata mentioned above in the destination-choice context relate to mode availability (children cannot drive alone, for example); thus they are important for the mode-choice log-sum calculation. However, the origin–destination mode utility itself (related to travel time and cost components) proved to be the same for each population strata. Thus, it proved to be more effective to store the origin–destination components of all 10 mode utilities instead of the 12 log-sum matrices. Thus, for each individual journey a log-sum subarray is calculated for 4,000 destinations and appropriate strata. Then, mode choice is implemented for the chosen destination without any significant additional time, because the log-sum (mode-choice formula denominator) and all 10 mode utilities are available at the moment for the record being processed.

It should be mentioned that applying the Monte Carlo random pick as well as the modification of the core choice model (such as using a nested logit model for mode choice instead of a multinomial logit model) does not impose any additional computational burden. The critical time-consuming part of the procedure relates to the manipulation (reading, calculation, storing) of large multidimensional utility and probability arrays.

The most important savings to achieve in modeling relate to reducing dimensionality whenever possible; microsimulation greatly helps with that objective. For example, essentially, the mode-choice model is applied for only one (the chosen) destination for each individual journey. In the conventional modeling framework, mode-choice probabilities should always be calculated for all destinations.

One possible way to reduce the size of the destination-choice set in application is to presample destination zones in a way similar to that done for the model calibration. However, statistical efficiency of sampling requires a priori importance to be assigned to each destination in conjunction with the given origin. Reasonable importance weights should incorporate basically the same attributes (impedances, size variables) as the destination choice model itself; thus, in computational terms, generating this sample set will likely add back the time saved on the model application.

Because of the easy and compact storage of microsimulation results, it is always possible to output a wide variety of reports across any travel dimensions or to single out any particular segment, for example, modal split for a specified person type only. In many cases, with conventional or standard disaggregate models, rerunning the model with changed reporting options would be required.

## CHAINED DECISIONS AND INDIVIDUAL TIME–SPACE CONSTRAINTS

### Behavioral Realism

Another attractive aspect of microsimulation is the way it can improve the behavioral realism of travel demand models. Explicitly modeling the behavior of individuals in households allows for the exploration of a chained or hierarchical structure of travel decisions, and the consideration of objective time–space constraints on an individual's daily-activity pattern. Almost any behavioral aspect that can be formulated and measured in relation to modeling dimensions can be introduced into the microsimulation modeling framework in a natural way. During microsimulation, these additions usually lead to only minor complications, whereas in the conventional modeling systems they can impose a significant complication.

This behavioral realism has been recognized and widely acknowledged as the primary theoretical advantage of microsimulation. However, experience with calibration and application of the NYBPM microsimulation model has shown that there is a certain price that must be paid before this benefit can be achieved. It involves an appropriate preparation of the core probabilistic model, including multiple conditioning and truncation of probabilities to incorporate additional linkages and constraints. If constraints are introduced in the application of the microsimulation model without appropriate adjustment of the core probabilistic model, serious biases can be produced no matter how reasonable the constraints are.

For example, microsimulation allows for linkages between one individual's home-based and non-home-based journeys; these linkages need to be accounted for in estimation. In the NYBPM it is assumed that the drive-alone mode is available for at-work journeys only if drive-alone was chosen for journey to work. However, to implement that constraint, the availability of the drive-alone mode must be considered during the at-work mode choice model calibration. Drive-alone must be excluded from the available set of modes in the estimation data for the at-work journey if it was not chosen for the corresponding journey to work.

If this probability conditioning is not carried out (in this case, if the at-work mode choice was calibrated with a full set of modes), a significant underestimation of the drive-alone mode will occur during application. Thus, introducing further behavior realism into the modeling system can produce significant systematic biases if the new linkages and constraints have not been properly supported by the probability conditioning of the core model. In general, microsimulation allows for the incorporation of complex behavioral constraints on individual travel of two principal types:

- Chained decisions, which require conditioning of modeled choices. In technical terms they are usually expressed in various Boolean variables reflecting previously made choices that should be equally used in both calibration and application. These indicators create a linkage across various choice models. In some cases this takes the form of hierarchical nesting, in which several choice dimensions are considered in a mutual model (usually, nested logit) and appropriate composite log-sum measures are carried from the lower levels to the upper levels of the hierarchy. In other cases, the linkage has a more complicated form that cannot be reduced to a single nested structure.
- Physical time–space constraints on the individual daily travel agenda, which require truncation of the choice probabilities across certain dimensions. In particular, a limited destination choice for several journeys made during the same day is important in conjunction with timing. If one of the journeys, typically, commuting to work, takes a great deal of time on top of the working-activity duration, this will naturally limit any other independent home-based journeys. Modeling time–space constraints in a simplified way can be done by means of discrete representation of time-of-day and space dimensions. Thus, the same Boolean technique can be used to link various choice models. This has been the working approach in most cases for the NYBPM. A more advanced approach that treats time and space as continuous variables could be applied (*1, 2*).

### Chained Decisions

Chained decisions in turn can be subdivided into the two following categories:

- Sequential choices made by the same person, and
- Intrahousehold interactions in which travel decisions made by one household member can affect travel decisions made by the other household members.

Sequential choice made by the same person can be linked in the hierarchical nested structures. In several reported microsimulation models, an individual's whole daily activity agenda has been modeled as one multilevel nested structure (*3, 4*). In the NYBPM we have found that an interlinked set of journey-generation models can be calibrated and applied better than a single nested structure.

Additionally, such a set of linked models can incorporate intrahousehold interaction among several household members. We have found that intrahousehold interactions are relevant for mutual journey making, especially for household maintenance and discretionary purposes. For mandatory purposes that are more individual in nature (work, school, university), mutual journey making is less relevant and takes mostly the form of carpooling with passenger pickups and drop-offs on the way.

Linkage of choices made by different individuals cannot be modeled by means of the nested structure because the nested structure in a rigorous behavioral sense is still a one-decision-maker model. However, a sort of quasi-nested hierarchical structure can emerge and be calibrated in cases in which different levels refer to different decision-making units. For example, upper levels can represent an entire household, whereas the lower levels can represent individual household members (*5*). In our view, quasi-nested constructs with differential decision-making units by levels should be further explored from a theoretical point of view because they constitute a powerful practical tool for consistent modeling of the travel behavior of individuals linked in the same household.

#### *Linkages of an Individual's Travel*

The first group of chained decisions relates to the linkage of travel choices made by the same person during a day. It starts with a sequential modeling of journeys made for different purposes at the stage of journey production (frequency). In the NYBPM, journey purposes are classified by three activity categories: mandatory, household maintenance, and discretionary. Making journeys in the primary category (mandatory relative to maintenance and discretionary, maintenance relative to discretionary) affects journey frequency choice of the inferior category. Boolean indicators of "other journey" making proved to be significant variables with negative coefficients in the journey-frequency utility expressions reflecting that, all else being equal, implementing journeys of the superior categories leaves less room in the daily activity agenda for journeys from the inferior categories.

Another recognized advantage of microsimulation relates to mode-choice modeling. First, mode choice is applied after destination choice, so the destination-related attributes are known at the mode-choice stage for each individual journey. Having access to this information allows for a virtually unlimited set of variables to be used, including any socioeconomic stratification of either person and household characteristics or destination-based parameters. In the conventional fraction-probability modeling framework, that would immediately lead to a multidimensional probability array of an infeasibly large size.

Second, microsimulation creates the natural opportunity to model chained mode choice in which modes for different legs of the same tour are chosen in a consistent way. As an example, in the NYBPM, mode choice for at-work journeys is conditioned on mode choice for the appropriate journey to work.

Third, microsimulation allows us to actually model intermediate stops on each journey leg. Microsimulation creates an opportunity to link stop-frequency and location and mode-choice decisions in a behaviorally realistic way. In the NYBPM system, both stop-frequency and stop-location models are applied after mode choice and they intensively use mode-related attributes determined in prior steps.

On the other hand, stop-making indicators can be fed back to the mode-choice stage. That reflects a two-directional linkage between mode-choice and stop-making decisions in which both directions are equally important and cannot be unambiguously ordered. In some cases, travelers make stops because they have a convenient mode (drive-alone or shared ride) that has been chosen for the final destination (stop-making choice is conditional on mode choice). In other cases, the traveler chooses a mode (drive-alone or shared ride) because of the stop-making perspective (thus, mode choice is conditional on stop-making choice). An interesting two-level mode-choice construct that can partially resolve this contradiction has been applied in the San Francisco microsimulation model (*3*). It has an upper-level principal mode-choice model that is applied for an entire tour before modeling stops. After stops are modeled, the tour is broken into elementary legs and a detailed mode-choice model is applied for each leg conditional on the chosen principal mode.

Another important dimension of travel demand modeling in which microsimulation can bring a new depth of analysis relates to journey timing. Modeling time-of-day choice has always been a weak point for conventional travel demand models. In the past most models were limited to either the entire day or to a single time period, typically the a.m. peak. Microsimulation offers the opportunity to explicitly model time-of-day choice in the full context of other travel decisions. Time-of-day choice can be modeled by the discrete representation of the time dimension as a set of alternative time periods. A discrete-choice model for a pair of journeys (outbound leg and return leg) can be applied in which each alternative constitutes a feasible pair of time periods (say, a.m. for a journey to work with p.m. for the corresponding return-home journey from work).

Modeling time-of-day choice after destination and mode choice allows the inclusion of all available personal, household-related, origin-related, destination-related, and OD-related attributes in the choice utilities. Implicit multidimensional stratification corresponding to such a set of attributes would have made any application of conventional fractional-probability models impossible. However, in the microsimulation framework, no significant additional burden is imposed in application because the core model has been calibrated. Any activity duration– and travel time–related components can be included in the utility to properly model an implicit trade-off between activity duration and travel time savings that influences the traveler's decisions to switch from period to period.

Although placing the time-of-day choice after destination and mode choice offers significant advantages, it imposes the problem of determining which time-of-day period should be used to specify the level-of-service variables for destination choice. The best solution would be to use the full mode-choice log-sum during all periods. However, that is computationally cumbersome. A simplified approach adopted for the current version of the NYBPM is to use a predetermined time of day for each journey purpose and leg (for example, outbound journeys for work and school are almost exclusively implemented in the a.m. period). Additionally, an iterative feedback from time-of-day choice to mode and destination choice can be considered as a part of global equilibrium.

It was recognized that from the theoretical point of view, however, the time dimension would be better modeled as a continuous variable rather than an artificially discrete one. The corresponding apparatus is currently applied for isolated departure choice models (*2*). Although certain technical details must still be resolved before these models can be included in a practical modeling framework, it is certainly a promising method, especially in view of the time–space constraining technique that will in the future replace several discrete choice travel dimensions.

It is important to understand the relative advantages and disadvantages of a set of journey-frequency models linked by purposes as applied in the NYBPM, compared with an alternative overarching daily-activity-agenda model (*4, 6*). A theoretical advantage of the overarching daily-activity-agenda model is that it can capture interactions across all activity types (journey purposes), including feedbacks from lower-level (discretionary, maintenance) activities to the upper level (maintenance, mandatory). In this structure an entire set of daily activities achieved by travel constitutes a choice alternative.

It is doubtful, however, that discrete choice models available for practical calibration, such as the multinomial or nested logit model, can capture the complicated and differential structure of similarities embedded in such a choice set. From that perspective a set of properly linked models by journey purposes (activity types) can probably achieve a better flexibility. This set can be thought of as a hierarchical nested structure in which journey-making indicators create nests for the lower-level activities. However, this structure is more complicated and flexible than a simple nested logit because it allows for the fractional distribution of lower-level alternatives among several nests, whereas in the ordinary nested structure each alternative can be included in only one nest.

One important enhancement relates to inclusion of accessibility measures. Their inclusion makes the journey-generation model sensitive to network improvements. Additionally, a set of linked journey-frequency models naturally covers the journey making of each household member with a special set of variables related to journeys made by other household members. Existing examples of the daily-activity-agenda models are mostly person based (*6, 7*).

### *Effects of Intrahousehold Interaction on Travel*

The second group of chained behaviors accommodated by microsimulation involves intrahousehold interaction among different household members. More than 30% of the journeys made in the New York–New Jersey metropolitan area involve two or more household members. The journeys can either be fully joint (like many maintenance and discretionary journeys) or include drop-offs and pickups of household members on the way to work or university.

Several modeling components in the NYBPM belong to this group. One of them is applied at the journey-generation stage. The journey-frequency choice utilities for several purposes and personal categories include Boolean indicators of journeys made by other household members.

Experience with the NYBPM has shown that it would be beneficial to explicitly model the mutual journey making of several household members. That is especially relevant for maintenance and discretionary purposes. It makes mode choice much more selective, especially for the shared-ride mode, which is otherwise modeled for each person separately without an explicit linkage in the household. The microsimulation framework creates an optimal environment for modeling mutual journeys.

In the current version of the NYBPM, mutual journey making is not modeled explicitly. However, several mutual journey-making indicators have been introduced in the mode-choice models for maintenance and discretionary purposes. In particular, Boolean indicators of two adult household members having a journey for the same purpose (either maintenance or discretionary) and an adult and a child having a journey for the same purpose proved to be strong explanatory variables favoring the shared-ride mode.

Figure 1 represents a combination of the two groups of chained decisions (sequential choices of the same person and intrahousehold interactions) relating to the journey-frequency model adopted for the NYBPM. There are three person types (worker, nonworking adult, and child) and six journey purposes. This finally yields 13 journey-frequency models. That takes into account the fact that children cannot implement journeys to work, at work, and to university and that nonworking adults cannot implement journeys to work and at work.

A set of the journey-frequency models is ordered and linked in the following way:

1. School for children;
2. School for workers, having a child-at-home Boolean indicator that has a negative effect on journey making;
3. University for workers, having a child-at-home Boolean indicator that has a negative effect on journey making;
4. School for nonworking adults, having a child-at-home Boolean indicator that has a negative effect on journey making;
5. University for nonworking adults, having a child-at-home Boolean indicator that has a negative effect on journey making;
6. Work, having indicators for child-at-home event and other mandatory journeys made by the same worker that have a negative effect on journey making;
7. At work for those workers that made a journey to work;
8. Maintenance for nonworking adults, having indicators for child-at-home event, other mandatory journeys made by the same nonworker (these have a negative effect on journey making), and journeys to work made by working household members (these have a positive effect, reflecting the fact that nonworking adults usually take care of shopping, children's chaperoning, and other household-maintenance needs for workers);
9. Maintenance for workers, having indicators of child-at-home, other mandatory journeys made by the same worker, and maintenance journeys made by nonworking household adults (these have a negative effect on journey making);
10. Maintenance for children, having indicators for school journeys made by the same child (these have a negative effect for journey making) as well as for maintenance journeys made by adult household members—working and nonworking (these have a positive effect on journey making, reflecting intrahousehold cooperation in mutual journey making)—and journeys at work made by workers (these have a negative effect on journey making, reflecting a maintenance purpose for majority of at work);

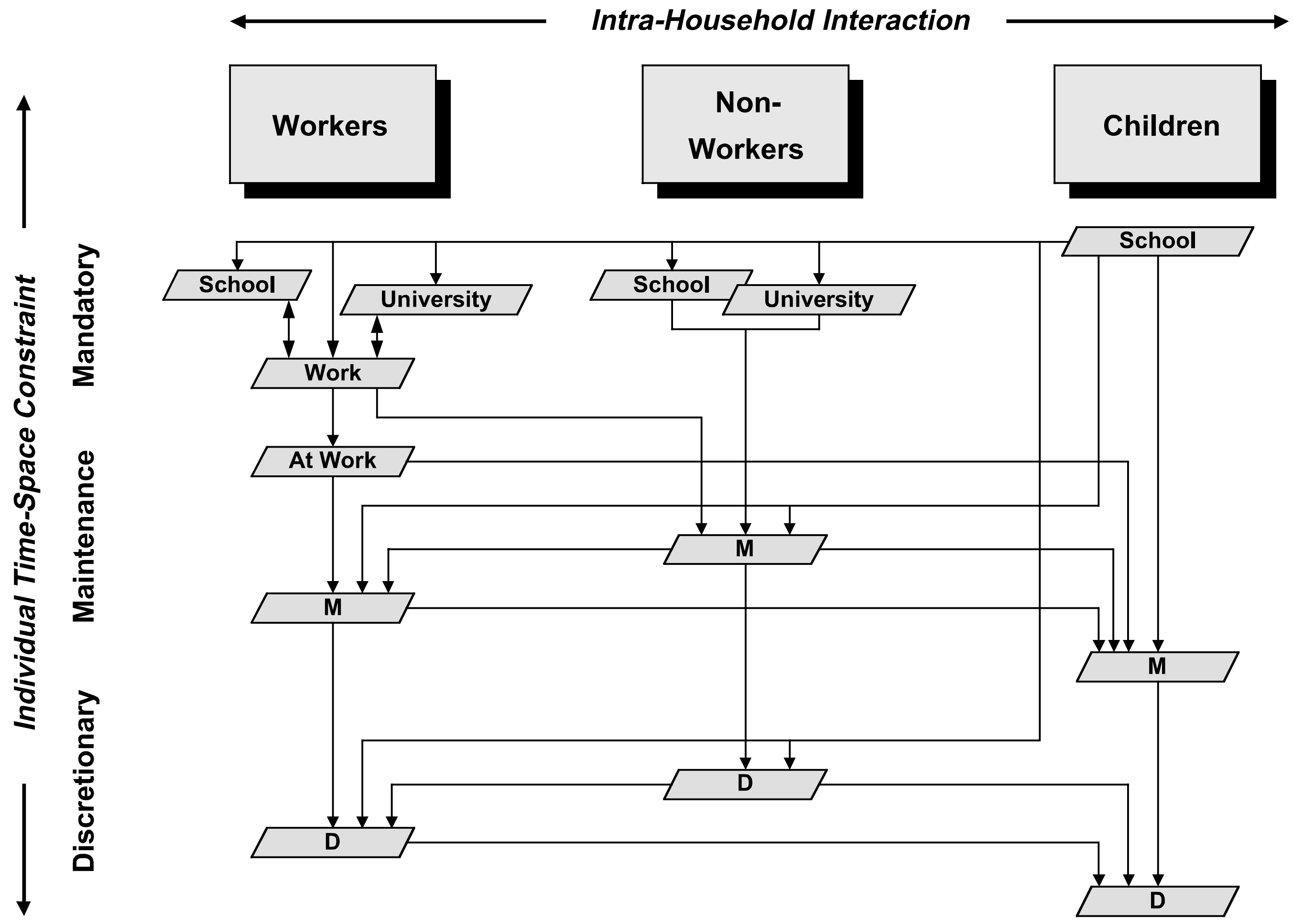


**FIGURE 1 Linked journey-frequency models.**

11. Discretionary for nonworking adults, having indicators for child-at-home event and other mandatory and maintenance journeys made by the same nonworker (these have a negative effect on journey making);

12. Discretionary for workers, having indicators for child-at-home event and other mandatory and maintenance journeys made by the same worker (these have a negative effect on journey making) as well as for discretionary journeys made by nonworking adults (these have a positive effect on journey making, reflecting intrahousehold joint journeys); and

13. Discretionary for children, having indicators for school and maintenance journeys made by the same child (these have a negative effect on journey making) as well as for discretionary journeys made by working and nonworking adults (these have a positive effect on journey making, reflecting intrahousehold joint journeys).

All indicators providing the linkage across household members or journey purposes proved to be highly significant statistically. Thus, both intrahousehold interaction (horizontal linkage) and individual time–space constraints (vertical linkage) have been recognized as important factors in household journey making.

## Time–Space Constraints

Explicit time–space constraints represent another technique to more realistically model each individual's daily travel agenda (*7–9*). To some extent these constraints are interchangeable with chained decisions. For example, alternatively, the negative effect of the Boolean journey-making indicators for primary activities (mandatory, maintenance relative to discretionary) in the journey-frequency utilities for secondary activities (maintenance relative to mandatory, discretionary) can be modeled by appropriately constraining individual travel in time and space during a day.

In the last case real physical constraints that reflect the sequence of destinations chosen by a person would definitely be better than simple Boolean indicators of journey making. An individual daily activity agenda can be meaningfully subdivided by several spatial–temporal "prisms." For example, commuting daily activity agenda can be naturally broken into before-work, at-work, and after-work prisms (*7, 8*). Then appropriate application of scheduling and sequencing rules in each prism can produce a sort of flexible nesting for daily activity and travel combinations.

The introduction of this attractive concept into practical demand modeling is hampered for a couple of reasons. First, journey-generation and spatial-distribution (destination choice) stages would merge into a single model that has not yet been developed. Second, the spatial dimension of real-world networks expressed in thousands of traffic zones adds a computational complexity that prohibits any choice models that do not have a simple analytical form (probit, generalized logit).

Microsimulation does offer certain new possibilities here, however. Dealing explicitly with each individual allows for a great reduction in the set of available zones for each journey and for the construction of a logical sequence of choices in a linked spatial dimension. In this sequence each consecutive choice has a narrowed set of available zones stemming from the time–space constraints and previous spatial choices.

Experience with NYBPM calibration supports certain practical rules for effective modeling of time–space constraints. For example, it has been found that residences and workplaces serve as important pivot points in space for commuters. All other destinations, either for stops during the commuting journey or for making independent home-based journeys, are located mainly in the ellipsoidal spatial envelope having these pivot points as focuses. That gives rise to a modeling structure in which workplace locations are modeled for household workers first. Then all journey-generation and spatial distribution models include commuting distance as an important explanatory variable that limits both the number of other journeys and the set of potential destinations (*3*).

A similar concept has been applied for the stop-location model in the NYBPM framework. It has been found that most stops occur in one of the three spatial domains adjacent to the journey pivot points:

- A 3- to 5-mi radius (depending on journey purpose) around the residential place;
- A 3- to 5-mi radius around the journey destination; or
- A middle region between the two locations, but with not more than a 20% to 50% deviation from the shortest path between them.

Thus, stop location is chosen from a reduced set of zones relevant for the OD pair of the corresponding journey. In future versions of the NYBPM, it is proposed that an upper bound be established for total individual daily mileage, which would be used for successive reduction of available destination zones for each modeled journey. Spatial-set-reduction strategies of this sort can be implemented only in the microsimulation framework. The conventional modeling framework would require an enormous additional stratification to handle time–space constraints.

The modeling rules described above, though practically effective, constitute only surrogates for a destination-choice (activity-location) model that ensures consistency for all journeys made by a person during a day. This modeling dimension is closely intertwined with the journey-timing and activity-duration dimensions. In the current version of the NYBPM, journey timing follows destination choice for each journey separately without explicit formulation of a daily schedule.

Two additional components could be introduced in future versions of the microsimulation-based NYBPM:

- A journey-sequencing model that would define the order of journeys in the course of a day, forming an individual's daily activity schedule; and
- A journey-timing model that would choose periods of the day for each journey leg, with the time dimension still broken into discrete periods.

Examples of corresponding modeling constructs can be found in the literature (*7–10*), although they have not yet been incorporated into a fully dimensional travel demand model. Microsimulation allows for the natural combination of spatial and temporal dimensions, especially if the journey-sequencing model is applied first; destination and time-of-day choice models are then applied according to the chosen sequence.

Another time–space constraint on household travel behavior involves car use and allocation among household members (*5*). In microsimulation, household cars can be modeled explicitly as an important travel resource. An additional model is required to allocate cars among adult household members by periods of a day and journeys. This would yield a significant modeling benefit—the real availability of an individual car, specific to each journey, instead of a mode choice based on household car ownership. With explicit tracking of car allocations among household members during a day, the mode-choice model can be reduced to a submode choice among only the transit modes in which the drive-alone and shared-ride modes have been modeled at the car-allocation stage.

## VARIABILITY OF MICROSIMULATION

Variability can be exploited as a significant advantage of microsimulation in certain circumstances, but it may be viewed as a problem in other cases. Ultimately, how it is viewed will depend on the sort of decision-making and overall planning environment in which a model is applied.

Estimating the distribution of expected traffic across a planned facility may be more important than simply predicting the average volume. Microsimulation provides a unique opportunity to explicitly model the expected range of likely traffic volumes. Thus, decisions on the capacity of a planned transportation facility can be made based on the probability of achieving critical maximum volumes, rather than on average daily or hourly volumes.

It is important to distinguish between pure microsimulation variability stemming from Monte Carlo realizations and systematic sources of travel demand variability (season, day of week, sports event, etc.). Observed systematic sources of variability should be included in the core model formulation, whereas Monte Carlo variation should account for various nonmodeled random factors on top of systematic fluctuations. Although seasonal and other systematic factors of variability can be modeled in the conventional framework, additional random factors remain hidden.

In the metropolitan-planning context, however, the variability of microsimulation results may be perceived by many practitioners as posing more drawbacks than advantages, particularly if the magnitude of the variability stemming from random processes proves to be large. Even after the unrealistic nature of apparently deterministic conventional models is acknowledged, variability may be viewed as a feature that confuses use of the models and interpretation of results. In discussions about how the microsimulation-based NYBPM should be applied, in conformity analysis, for example, a perceived need for an averaging guidance emerged. The client desired an application protocol that could be used to manage variability and bring the microsimulation model in line with the conventional planning framework. The simplest averaging strategy form is to conduct a predetermined number of microsimulation runs with different seeds and to mechanically average the results.

In general, if the result of applying the whole core probabilistic model is considered as a multidimensional distribution across various travel dimensions, each microsimulation run is a particular point (realization) in this multidimensional space. The sharper the core distribution and the more runs implemented, the higher is the probability that average statistics will replicate their probabilistic means. However, if the core distribution has substantial variance and the number of runs is limited, there is a higher probability that microsimulation will generate some extreme realization that can be far from the true average value.

Although the variability of microsimulation can be explored statistically from the theoretical point of view, the complexity of the core probabilistic model makes certain aspects of this analysis still partially empirical. On the basis of this analysis, rules can be established for application of microsimulation models, depending on the dimensionality of the model and planning purposes. The ultimate use of the model will dictate the desired number of runs with different seeds needed to produce a distribution of forecast volumes within the necessary degree of accuracy.

The limited framework of the current paper does not allow for further elaboration of this important aspect of microsimulation. The project team has explored microsimulation variability, both statistically and in relation to its important implications for planning. Certain rules for the estimation of variability in microsimulation have been formulated and then examined empirically using the NYBPM microsimulation system. The corresponding documentation is available from the authors.

## MANAGING COMPETITION IN MICROSIMULATION

### Constraining Travel Demand Models

Destination choice is one of the most complicated travel demand dimensions for two reasons. First, there are thousands of alternative zones available. The resulting sparseness and "lumpiness" of observed data make it very difficult to calibrate a destination choice with any precision. Second, there is a discrepancy between aggregation levels for productions and attractions. Journey makers are not heading to destination zones but rather specific attractions contained in zones. This discrepancy is not immediately apparent in conventional models, but it is obvious in microsimulation.

Because of the complexity, many models apply destination choice in a singly or doubly constrained fashion. Some researchers believe that constraining models is not an appropriate technique. However, the real world is resource constrained. Not everyone who wants to work in the central business district is able to. Furthermore, some jobs are in locations with poor accessibility; in a modeling framework without constraints, such attractions would probably not be filled.

As soon as constraints are introduced into the modeling framework, a rudimentary competition arises, which generally is ignored in standard models. Microsimulation has the potential to handle this competition over work attractions and other activities in a meaningful fashion. Introducing competition is not considered a primary goal of microsimulation, but rather a by-product of processing individual records sequentially in a constrained destination choice module. Microsimulation processing requires that enough attractions be made available for journey makers or the modeled productions cannot be matched and the journeys will not be completed.

There are significant differences in how to apply constraints in a microsimulation setting compared with a conventional model. In conventional models in which destination choice is doubly constrained, aggregate flows between zones are calculated using some variant of the gravity model approach. Essentially, competition still exists in these models, but it is hidden inside balancing factors. Applying constraints in this manner is particularly inappropriate for microsimulation because discrete journeys are replaced with fractional journeys, breaking the link between the journey maker and the journey.

To retain discrete journeys, a new type of constraining technique was developed for NYBPM. Journey makers are sequentially assigned to specific journey attractions and, more important, for a constrained purpose; once selected, these attractions are removed so that another journey maker cannot choose that particular attraction. Probabilities of choosing various destinations are periodically updated to account for those attractions already exhausted.

### Technical Application of Destination Choice Constraints in Microsimulation

To apply the destination choice model in microsimulation, the purpose-specific zone attractions are generated. Balancing is then carried out by proportionally adjusting total attractions to match regional productions. Because every production must be assigned to a specific attraction, the destination-choice module will fill out a production-attraction (PA) matrix. If the purpose is fully constrained, every single attraction will be chosen, and the resulting matrix will automatically be balanced. At this stage it is necessary to declare the constraint parameter for that purpose, which is used to inflate the number of available attractions. For a fully constrained purpose such as work, school, or university journeys, the constraint factor would be 1. For a relaxed constrained purpose, any value greater than 1 will increase attractions and loosen the constraints. In the NYBPM, maintenance and discretionary journeys were relaxed constrained with a scaling factor of 1.5 to 2. Using a value of 99999 will result in an unconstrained model.

Transportation researchers have considered relaxed constrained modeling a potentially attractive, but elusive option for quite some time because of the practical difficulties of calibrating and applying relaxed constraint destination choice models (*11*). In contrast, switching a destination choice model from full constraints to relaxed constraints in microsimulation is very straightforward.

As each journey record is assigned to a destination, the stock of attractions in the chosen zone decreases by 1. The effective rate of decrease is controlled by inflating attractions using the constraint parameter. As attractions are removed, the zone probabilities must all be recalculated. The probability of choosing this zone decreases slightly with each realization, and zones that have not lost attractions gradually become more attractive. When all attractions in a particular zone are taken, the destination zone is closed for that purpose.

A significant computational effort is required to recalculate and store these probabilities after every single journey is assigned. Thus, a packeting strategy that processed similar records together was developed. It substantially cut down on computer time; the attractions are recalculated at selected intervals in the processing (typically, every 10 or 20 records), rather than after every single record. In bundling households with similar characteristics together for processing, NYBPM is following an accepted microsimulation technique (*12, 13*). The strategy does introduce slight discrepancies in the destination-choice probabilities. Because zones cannot close until all records in a packet have been processed and the probabilities are refreshed, it is possible to slightly overassign journeys to a zone. Thus, there is a trade-off between precision in constraining the PA matrix and saving computer time.

It is the manner of closing zones in the mode and destination choice model that introduces explicit competition over jobs and, to a lesser extent, maintenance and discretionary activities. At some point the most attractive zones—those with reasonable accessibilities and many employment opportunities—will be filled. The destination choice component will then force journey makers to travel to less attractive zones. Toward the end of the processing of journey records, simulated workers face a landscape in which only a few zones remain open and they will be assigned to the remaining work attractions regardless of accessibility. Nearly all the long journeys in the tail of the journey length distribution are the records processed last. This is the so-called last-record problem.

### Empirical Research on the Last-Record Problem

Some form of the last-record problem will arise in any modeling procedure in which records are processed sequentially and constraints are employed. This last-record problem can also introduce some discrepancy with the calibrated parameters in destination choice because, typically, estimation is carried out as if all journey makers could travel to all opportunities.

Medium-income-work journeys display the same overall pattern as do the other fully constrained purposes. In general, the smaller the purpose, as measured by journey productions, the more affected it will be by the last-record problem. There is a sharp increase in the average distance of motorized medium-income-work journeys near the 80% mark. Motorized journey makers in the last slice, on average, travel more than 30 mi to work compared with 10 mi if they are in the first 75% of records processed (see Figure 2).

In contrast, the last-record problem does not significantly affect relaxed constrained purposes. Average journey distance for both motorized and nonmotorized travel looks constant for all records. Motorized journeys do not depart from an 8-mi average distance. This indicates that enough destinations remain open for journey makers to be successfully placed all the way through the record file. The results for discretionary journeys were almost identical to those for maintenance.

Because applying microsimulation with constraints introduces the last-record problem, modelers must develop techniques to ensure reasonable results. If too many of the journeys are extremely long because of destination-choice constraints and the resulting competition, the situation must be addressed. At the same time, at least some long journeys should be generated, because commuters in most metropolitan regions do make very long journeys, as revealed in journey-length frequency distributions estimated from household travel surveys and the journey-to-work data from the census.

The first step is to compare the journey length distribution model results against available household survey data. In general calibration, achieving the proper mix of short- and medium-length journeys is most important. In this case, however, because we are investigating the last-record problem, we will focus exclusively on very long journeys (75+ mi) to see whether the tails are comparable.

In the NYBPM, results vary by travel purpose. Although all purposes produce at least a few very long journeys in application, only the fully constrained purposes generate enough journeys in the tail (0.3% or more) to meaningfully affect average journey length, which is consistent with our expectations (see Table 1). In fact, the relaxed constraint purposes do not appear to be producing enough very long journeys, which suggests that the constraint factor could be tightened up. Within the fully constrained purposes, medium- and high-income work appear to be in line with the survey targets for very long journeys, but low-income work and university journeys appear to be adversely affected by the last-record problem.

As a general principle, purposes with major spatial imbalances between productions and attractions will be the most affected by the last-record problem. It might also be expected that relatively small purposes would be affected most, simply because there are so few productions and attractions to work with. In a model with nearly 4,000 zones (16 million cells), all PA matrices will be sparse, but purposes below 1 million journeys are extremely sparse. Thus, it is not surprising to see that low-income-work journeys and university journeys (small, fully constrained journey purposes with inherent spatial imbalances) are the purposes most affected by the last-record problem.

**TABLE 1 Long Journeys—Survey Versus Model**

| Travel purpose | Proportion of long journeys (75+ miles) | | Purpose size (millions of journeys) |
|---|---|---|---|
| | Survey data (%) | Model output (%) | |
| Fully constrained, globally balanced purposes | | | |
| Low-income work | 0.0 | 1.1 | 0.3 |
| Medium-income work | 0.7 | 0.3 | 5.0 |
| High-income work | 0.3 | 0.3 | 1.6 |
| University | 0.5 | 1.3 | 0.5 |
| Fully constrained, subregionally balanced purpose | | | |
| School | 0.5 | 0.0 | 3.1 |
| Relaxed constrained, globally balanced purposes | | | |
| Maintenance | 0.6 | 0.0 | 8.5 |
| Discretionary | 0.5 | 0.0 | 2.6 |

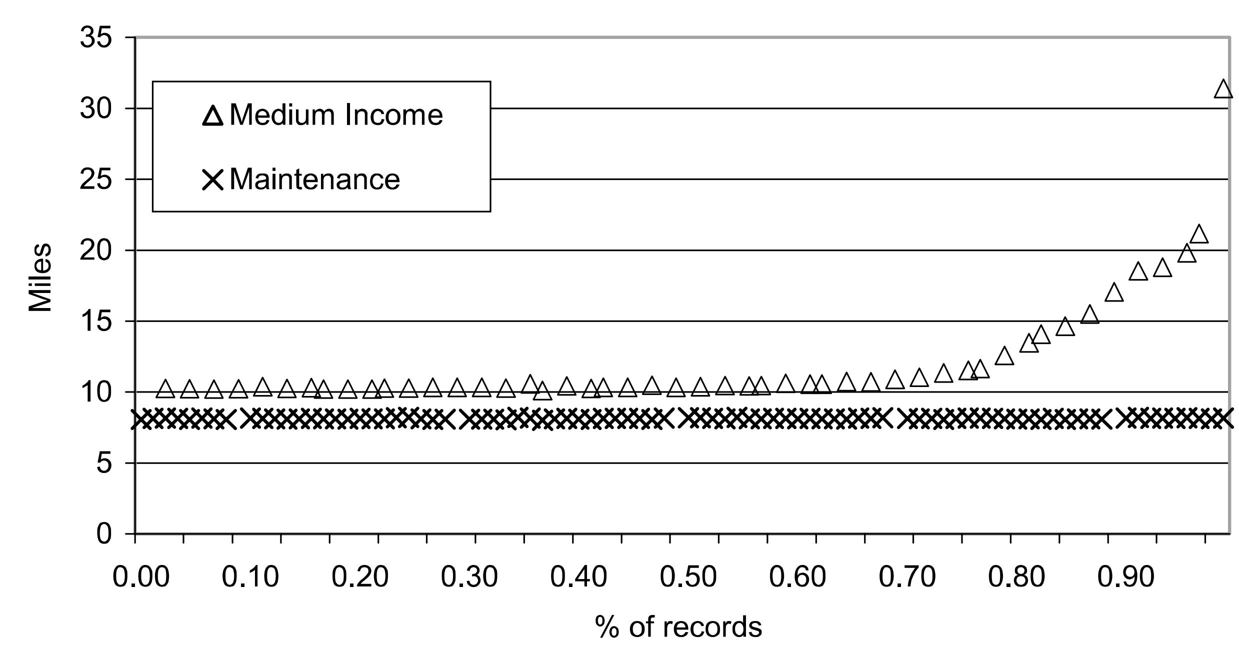


**FIGURE 2 Average motorized trip distance.**

The current implementation of the NYBPM applies a variety of techniques to minimize the last-record problem and to keep destination-choice results reasonable. First, the synthesized households are grouped into packets that share the same origin zone and household characteristics. Then these packets are processed in a random order. Although this random order approach cannot guarantee a reasonable allocation of work journeys, nor will it completely eliminate the last-record problem, it will prevent systematic biases from creeping in. We believe that more research will have to be done to handle competition whenever models are constrained during destination choice.

## CONCLUSIONS

Several conclusions can be made:

- Microsimulation is not an alternative modeling approach to the conventional demand models; it is a natural extension of the disaggregate modeling technique in which simulating trip makers on an individual basis allows for complex linkages across multiple trips, ultimately resulting in a better estimation of real-world travel behavior.
- Microsimulation has three major advantages over conventional travel models: (*a*) substantial savings in the calculation of multidimensional fractional-probability arrays, (*b*) explicit formulation of various chained decisions and time–space constraints on individual travel behavior, and (*c*) explicit modeling variability of travel demand rather than average values.
- Microsimulation offers substantial savings in the calculation and storage of multidimensional fractional-probability arrays during model application. It opens an easy way to output a wide variety of reports by any combination of travel dimensions without rerunning the model. This is especially relevant for large metropolitan areas.
- Microsimulation allows modelers to improve the behavioral realism of travel demand models. Explicitly modeling individuals in households allows for exploration of a chained or hierarchical structure of travel decisions as well as objective time–space constraints on a daily-activity agenda.
- To take full advantage of microsimulation, the core probabilistic model must be restructured to take into account chained travel behavior. Although the final results will be more realistic than unlinked decision making, the estimation stage requires more effort because of the appropriate complication of the core probabilistic model.
- Microsimulation provides a unique opportunity to explicitly model the variability of transportation flows. Thus, a decision on the capacity of the planned transportation facility can be made on the basis of the probability of achieving critical maximum volumes, rather than on simple average daily or hourly volumes.
- When constraints are introduced into the modeling framework at the destination choice stage, competition over attractions arises. Generally, this competition has been ignored in traditional travel demand models. Microsimulation has the potential to handle the competition over work attractions and other travel activities in a meaningful fashion, though more research is needed to meet that goal.
- Constraining microsimulation models leads to the last-record problem because records are processed sequentially. Purposes with major imbalances between productions and attractions will be the most affected by the last-record problem, particularly when the purposes are also fully constrained. One promising line of research involves determining how to set the constraint factors for relaxed constraint purposes.

## ACKNOWLEDGMENTS

The authors wish to acknowledge the significant contributions to this research made by Gordon Schultz, Jim Hicks (PB Consult), Jeff Bruggeman, William Woodford (AECOM), and Bill Allen. The NYBPM has been developed for the New York Metropolitan Transportation Council, and the authors thank Tom Schulze and all of his technical staff for their support and guidance in this effort. In addition, the authors would like to thank the anonymous referees, who provided several valuable suggestions.

## REFERENCES

1. Bhat, C. R., and F. S. Koppelman. Retrospective and Prospective Survey of Time-Use Research. Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C., 1999.
2. Bhat, C. R., and J. Steed. Continuous-Time Model of Departure Time Choice for Urban Shopping Trips. Presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.
3. 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. Presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.
4. Bradley, M., M. L. Outwater, N. Jonnalagadda, and E. R. Ruiter. Estimation of Activity-Based Microsimulation Model for San Francisco. Presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.
5. Wen, C.-H., and F. S. Koppelman. Integrated Model System of Stop Generation and Tour Formation for Analysis of Activity and Travel Patterns. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1676,* TRB, National Research Council, Washington, D.C., 1999, pp. 136–144.
6. Bowman, J. L, and M. E. Ben-Akiva. Day Activity Schedule Approach to Travel Demand Analysis. Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C., 1999.
7. Gangrade, S., and R. M. Pendyala. Nested Logit Model of Commuters' Activity Schedules. Presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.
8. Yamamoto, T., R. Kitamura, and K. Kishizawa. Markov Chain Monte Carlo Algorithm for Generating Activity-Travel Patterns. Presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.
9. Bhat, C. R., and F. S. Koppelman. Activity-Based Travel Demand Analysis: History, Results and Future Directions. Presented at the 79th Annual Meeting of the Transportation Research Board, Washington, D.C., 2000.
10. Bhat, C. R. Duration Modeling: A Methodological Review with Applications in Activity-Travel Analysis. Presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.
11. Fotheringham, A., and M. O'Kelly. *Spatial Interaction Models: Formulations and Applications.* Kluwer Academic Publishers, Dordrecht, Netherlands, 1989.
12. Brog, W., and D. Zumkeller. An Individual Behavioral Model Based on the Situational Approach (Sindividual). In *Research for Transport Policies in a Changing World, Proc., World Conference on Transport Research,* Hamburg, Germany, Vol. 1, 1983, pp. 90–106.
13. Fellendorf, M., and P. Vortisch. Integrated Modeling of Transport Demand, Route Choice, Traffic Flow, and Traffic Emissions. Presented at the 79th Annual Meeting of the Transportation Research Board, Washington, D.C., 2000.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.*