Disaggregate Attraction-End Choice Modeling
Formulation and Empirical Analysis

CHANDRA BHAT, AJAY GOVINDARAJAN, AND VAMSI PULUGURTA

For travel demand models to provide good forecasts, they must be causal; that is, the models should represent the travel decisions made by individuals (and households) and should incorporate important demographic and policy-sensitive explanatory variables. This recognition has led to a shift from the aggregate modeling paradigm to the disaggregate modeling paradigm, evident in the widespread use of disaggregate trip production and mode choice models in practice. However, this shift toward disaggregate procedures has not yet influenced the fundamental specification of trip attraction and distribution models employed in practice. Developed and estimated were disaggregate attraction-end choice models that will facilitate the replacement of the aggregate trip attraction and distribution models currently in use. The proposed disaggregate attraction-end choice model is compared with the disaggregate equivalent of the gravity model.

For demand models to reflect accurately traffic flow changes in response to potential changes in the transportation infrastructure or population sociodemographics, they must be causal; that is, the travel demand models should represent the travel decisions made by individuals (and households) and should incorporate important demographic and policy-sensitive explanatory variables (*1*,*2*). This recognition has led to a shift in the past couple of decades from the aggregate (or zonal level) modeling paradigm toward the disaggregate (or individual-household level) modeling paradigm, as evident in the widespread use of disaggregate trip production and mode choice models in practice (*3*–*5*). However, the shift toward disaggregate procedures has yet to influence the fundamental specification of trip attraction-end and trip distribution models (*6*).

The objective of this research was to formulate and estimate (disaggregate) attraction-end choice models that will facilitate the replacement of aggregate trip attraction models and the aggregate trip distribution models currently used by most metropolitan planning organizations. The traditional planning process first estimates trip productions from each zone and trip attractions to each zone and subsequently links trip productions to trip attractions in the trip distribution stage. This process will be revised with the introduction of attraction-end choice models. In this revised procedure, trip productions from each zone first are estimated by using disaggregate methods. Subsequently, the attraction-end of each trip production is determined by the attraction-end choice model, and this provides the trip interchanges between each zonal pair. The total trip attraction to each zone is determined implicitly because it is simply the summation of all paired trip interchanges involving the relevant zone as the attraction-end. As part of the objective to estimate disaggregate attraction-end choice models, the research also evaluated the performance of the disaggregate choice models with the disaggregate equivalent of the conventional gravity model.

C. Bhat, University of Texas, Austin, TX 78712. A. Govindarajan, University of Massachusetts, Amherst, MA 01003. V. Pulugurta, KPMG Peat Marwick, 8200 Greensboro Drive, Suite 400, McLean, VA 22102-3803.

DISAGGREGATE ATTRACTION-END CHOICE MODEL FORMULATION

Attraction-End Choice Alternatives

The proposed disaggregate attraction-end choice model predicts the individual choice of travel to aggregated zones (or spatial clusters) and not to specific attractions within zones (or elemental alternatives).

The choice of trip attraction-end is characterized by a large number of alternatives even after defining alternatives at the zonal level. However, by adopting an identically and independently distributed (IID) structure for the error terms across the zonal attraction alternatives, the attraction-end choice model can be consistently estimated with only a subset of alternatives in the choice set (*7*). The subset of alternatives can be drawn in many different ways from the feasible choice set (*8*). The simplest method is to use a random sampling approach in which the subset includes the chosen attraction-end alternative and a random sample of nonchosen alternatives. This method was adopted in the current research. The total number of attraction-end alternatives sampled for each trip production is prespecified to be seven (six randomly selected alternatives and the chosen alternative).

Model Formulation

The alternatives in the attraction-end choice models are aggregate zones. Each zone j may include several possible elemental attraction alternatives. Let the number of elemental attractions in zone j be D_j. Assume the following: (*a*) D_j is large for each zone j, (*b*) utilities of the elemental alternatives within each zone are IID (conditional on unobserved zonal attributes), and (*c*) the systematic utility of the elemental attractions are relatively homogenous within each aggregate zone or the within-zone variance of the systematic utilities of the elemental alternatives are about equal across zones. With the foregoing assumptions, the utility (or benefit) U_{ij} presented by attraction-end zone j for a trip production made by an individual q in zone i may be written as (*9*):

$$U_{ijq} = V_{ijq} + \epsilon_{ijq} = \mu' Z_{ijq} + \eta \log D_j + \epsilon_{ijq} \quad (1)$$

where Z_{ijq} is a vector comprising (*a*) travel impedance measures for travel between zones i and j, (*b*) zonal attractiveness and location attributes of candidate attraction zone j, and (*c*) interactions of the sociodemographic characters of individual q with travel impedance and zonal attributes. D_j is the number of elemental alternatives in zone j, ϵ_{ijq} is a random term distributed IID gumbel across zonal alternatives and individuals, and μ and η are parameter vectors to be estimated. In Equation 1, D_j (the number of elemental work or shop attractions in zone j) is not easily quantifiable. However, we can proxy D_j by a set of observable size variables such as employment in zone j or land area of zone j. Let d_j represent a vector of proxy size variables for zone j and let δ be a corresponding vector reflecting the contribution of the proxy size variables to the actual zone size D_j. Then Equation 1 can be rewritten as

$$U_{ijq} = \mu' Z_{ijq} + \eta \log\left(\delta' d_j\right) + \epsilon_{ijq} \tag{2}$$

The magnitude of the parameter η in Equation 2 characterizes the presence of common unobserved zonal attributes (such as congestion levels or parking availability) affecting the attractiveness of elemental alternatives in a zone (*9,10*).

The probability that individual q will choose destination zone j from the set C_i of attraction-end zones accessible from zone i may be formulated as

$$P_{ijq} = \frac{e^{V_{ijq}}}{\sum_{j' \in C_i} e^{V_{ij'q}}} \tag{3}$$

The model specified in Equation 3 may appear to be the familiar multinomial logit model. However, this is not the case: the expression in Equation 3 includes the component $\eta\log(\delta' d_j)$. Thus, the model takes the form of a nonlinear-in-parameters multinomial logit model.

The estimation of the model represented by Equations 2 and 3 is accomplished by using the maximum likelihood method. Maximization of the log-likelihood function is achieved by using the GAUSS matrix programming language.

DATA SOURCE AND SAMPLE USED

The data source for this study was a household activity survey conducted by the Central Transportation Planning Staff (CTPS) of the Boston Metropolitan Organization in April of 1991. The survey collected data on sociodemographics of the household and each individual in the household. The survey also included a one-day (midweek working day) activity diary to be filled out by all members of the household more than 5 years old. The sociodemographic information and the activity information were assembled in a person/household sociodemographic file and a trip file, respectively, by the CTPS.

The Boston metropolitan area is divided into 986 traffic zones. CTPS provided a zonal demographic and land-use file containing zonal attribute data for each of the 986 traffic zones. CTPS also provided a zone-to-zone travel impedance file for travel between each zone pair in the region.

The sample used for this analysis was generated from the survey trip file, person/household sociodemographics file, zonal demographics and land-use file, and zone-to-zone impedance file. The steps in the sample formation process are as follows (*11*):

1. Home-based work and home-based shopping trips were selected from the trip file and the origin-destination ends of each trip were converted to production-attraction ends.
2. The travel survey trip file was matched with the person/household sociodemographic file to append the sociodemographic characteristics of the individual to each of her or his trip productions.
3. A descriptive analysis was conducted on trip length to obtain the longest trip for each of the home-based work and home-based shopping purposes. The longest trip length was about 116 km (72 mi) for both trip purposes, and this distance was selected as the basis for determining the feasible attraction choice set of zones for each production zone.
4. The feasible choice set was identified for each production zone; six candidate attraction zones (but not the actual chosen attraction zone) were selected randomly from the feasible choice set for each trip production and the actual chosen attraction zone was added to these six randomly chosen alternatives (for a total of seven attraction-end alternatives).
5. Each trip production in the trip file was replicated seven times and each attraction-end alternative generated in the previous step was appended to one of the seven replicated records of each trip production.
6. The impedance values by all available modes for the production-attraction zone pair of each record in the replicated trip file were appended from the zone-to-zone impedance file; the impedance values were available by four time periods in the day and the values corresponding to the time-of-day of each trip productions was selected for appending.
7. The zonal attributes of each candidate attraction end alternative were appended from the zonal demographics and land-use file.

The final sample files included 7,337 trip cases ($7{,}337 \times 7 = 51{,}359$ records) for the home-based work purpose and 7,963 trip cases ($7{,}963 \times 7 = 55{,}741$ records) for the home-based shopping purpose.

VARIABLE SPECIFICATION

This section discusses the explanatory variables considered in the analysis and the functional form used for them. Six sets of explanatory variables were considered for inclusion in the attraction-end choice models: (*a*) impedance variables, (*b*) zonal size measures, (*c*) zonal attractiveness measures, (*d*) zonal location indicators, (*e*) a zonal spatial structure measure, and (*f*) interaction of sociodemographic variables with impedance and zone-associated variables.

Impedance Variables

The impedance variables include in-vehicle travel time (IVTT), out-of-vehicle travel time (OVTT), and travel cost (COST). The impedance measures were obtained from the CTPS for each zonal pair and by four time periods of the day. The appropriate impedance measures were appended to each production zone—candidate attraction zone pair based on the time period of the individual trip. The impedance variables were defined for each available mode between the production zone and the candidate attraction zone and were undefined for unavailable modes.

Three issues were addressed in the introduction of the impedance variables in the attraction-end choice modeling. First, substantial multicollinearity was found among different impedance measures (IVTT, OVTT, and COST) for each travel mode. Second, differential modal availabilities for different zone pairs required consideration. Third, a functional form for the introduction of the impedance effect in the choice modeling had to be developed.

The issue of collinearity in IVTT, OVTT, and COST was resolved by converting OVTT and COST into equivalent IVTT units for each available mode. The universal set of modes included highway, transit, and walk (the bike mode accounted for very few trips for both work and shopping purposes and so is not considered here). The equivalent IVTT units for each available mode (for each zone pair) was computed as follows:

$$C \text{ (equivalent highway IVTT units)} = \text{Highway IVTT} + \theta * \text{Highway OVTT} + \lambda * \text{Highway Cost}$$

$$T \text{ (equivalent transit IVTT units)} = \text{Transit IVTT} + \theta * \text{Transit OVTT} + \lambda * \text{Transit Cost}$$

$$W \text{ (walk impedance)} = \Delta * \text{Walk Time} \tag{4}$$

In this expression, values for θ and λ were obtained from work mode choice modeling among motorized modes. The ratio of the parameters for out-of-vehicle time and in-vehicle time from work mode choice estimation was 1.75 and this value was applied for θ. The money value of in-vehicle travel time was found to be $4 per hour in the mode choice modeling. This translates to an equivalent of 0.15 min of IVTT for each cent of travel cost, which is the value employed for λ (travel times are measured in minutes and cost in cents in the data). The value of Δ should be between 1 and 1.75. If $\Delta = 1$, this implies that the disutility of 1 min of walk time is the same as 1 min of in-vehicle time on motorized modes. If $\Delta = 1.75$, the disutility of walk time is considered to be the same as OVTT on motorized modes. The appropriate value for Δ was determined through empirical estimation.

The second issue in the introduction of the impedance variables related to the differential modal availability among zone pairs. The highway mode is available for all zone pairs in the Boston metropolitan area. However, the transit and walk modes are not available universally. Consider a production zone *i* and two candidate attraction-end zones *j* and *k*. Let the highway impedances C_{ij} and C_{ik} (in equivalent IVTT units as computed from Equation 4) be the same and let zone *k* be served by only a subset of the modes that serve zone *j*. Then the overall impedance assigned for travel from zone *i* to zone *j* should be lower than that assigned for travel from zone *i* to zone *k*. This fundamental principle may be accommodated by specifying a composite impedance for travel using a parallel conductance formula. To develop this formulation, define a transit availability dummy variable y_t (for each zone pair) that takes a value of 1 if transit is available for travel between the zone pair and 0 otherwise. Also define a walk availability dummy variable y_w that takes a value of 1 if walk is available and 0 otherwise. The composite impedance *H* is then written as

$$H = (1-y_t)(1-y_w)C + y_t(1-y_w)\left\{\frac{C}{1+\frac{C}{T^\beta}}\right\} + y_w(1-y_t)\left\{\frac{C}{1+\frac{C}{W^\gamma}}\right\} + y_t y_w\left\{\frac{C}{1+\frac{C}{T^\beta}+\frac{C}{W^\gamma}}\right\} \tag{5}$$

If both transit and walk are unavailable, the first term applies and the composite impedance is equal to the highway impedance *C*. If transit is available, but walk is not, the second term applies. If walk is available, but transit is not, the third term applies. If all modes are available, the fourth term applies. β and γ are positive parameters that indicate the relative weights placed on transit and walk modes, respectively, relative to highway as the mode of travel. If $\beta > 1$ ($\beta < 1$) then the highway mode dictates attraction-end choice more (less) than the transit mode (when both modes are available). Similarly, if $\gamma > 1$ ($\gamma < 1$) then the highway mode determines destination choice more (less) than the walk mode (when both modes are available). The relative importance between the transit and walk modes can be determined by comparing the values of β and γ. The transit mode determines choice more or less than the walk mode depending on whether $\beta < \gamma$ or $\beta > \gamma$.

The third issue in the introduction of travel impedance is the functional form for the effect of composite impedance on utility. The two forms considered were the linear form and the log-linear form, both of which have been used in earlier studies (*12*). The linear form implies that the marginal deterrence due to travel impedance is independent of the existing impedance level, and the log-linear form implies that the marginal deterrence decreases as the existing travel impedance level increases (i.e., a constant increase in the composite impedance has a higher deterrence when the initial impedance level is low than when the initial impedance level is high).

The determination of the value for Δ (related to the first issue) and the values for β and γ (associated with the second issue), and the functional form for the composite impedance term (associated with the third issue), was based on empirical estimation. The probability that a trip production from zone *i* will be attracted to zone *j* is given by the following nonlinear-in-parameters multinomial logit model (ignoring the effect of other nonimpedance measures):

$$P_{ij} = \frac{e^{\alpha f(H_{ij})}}{\sum_k e^{\alpha f(H_{ik})}} \tag{6}$$

where $f(H_{ij}) = H_{ij}$ for the linear functional form and $f(H_{ij}) = \ln(H_{ij})$ for the log-linear functional form. H_{ij} is the composite impedance as given by Equation 5 for travel from zone *i* to zone *j* and α is the coefficient on the composite impedance term. The model in Equation 6 was estimated with a specialized maximum likelihood code written in the GAUSS matrix programming language using the individual choice sample assembled for the attraction-end modeling. The model was estimated for three predefined values of Δ 0: 1, 1.35, and 1.75. For both the linear and log-linear forms, little sensitivity was found in the log-likelihood function at convergence or in the estimated values for β and γ to the value of Δ used. The best results, however, were, obtained when $\Delta = 1.00$ for both the linear and the log-linear form. Between the linear and log-linear functional forms, the log-linear form was found to perform substantially better. The final results of the estimation in this log-linear specification were as follows: $\beta = 1.0752$ and $\gamma = 0.8779$ for the work purpose and $\beta = 1.6155$ and $\gamma = 0.9988$ for the shopping purpose. All these estimates (except γ for the shopping purpose) were highly significant in their difference from the value of 1. These results indicate that, for both the work and shopping purposes, the highway mode determines attraction-end choice more than the transit mode, but the walk mode determines choice more than the highway and transit modes. However, for the shopping purpose, the difference in the effects of the highway and walk modes are not statistically different. By using the estimates of β, γ 0, and Δ, a composite impedance value (in equivalent highway in-vehicle time units) was computed using Equation 5 and introduced in a log-linear form along with other exogenous variables discussed below.

Zonal Size Measures

The zonal size measures considered in the analysis included total zonal employment for the home-based work purpose and retail plus service employment and zonal area for the home-based shopping purpose. The size variables represent proxy measures of the number of elemental destinations within a zone and so the expectation is that large zones more likely will be chosen as the attraction-end than small zones.

Zonal Attractiveness Measures

The zonal attractiveness measures included percentage unemployment rate and crime rate for the home-based work purpose, and crime rate for the home-based shopping purpose. The percentage unemployment rate was obtained from the Department of Employment and Training (*13*). The unemployment rate represents the percentage of individuals in the labor force residing in the zone who are not employed. Crime rate information was obtained from a report by the Massachusetts State Police Crime Reporting Unit (*14*). Crime rate was measured as the total number of police-reported crimes for 1992.

Zonal Location Indicators

The zonal location indicators represent the geographic location of a zone relative to the location of the Boston central business district (CBD). The CTPS developed a ring system of geographic location. Ring 0 includes the zones in and immediately around the Boston CBD. Ring 4 encompasses the rural zones in the periphery of the Boston metropolitan region. Rings 1, 2, and 3 lie in between. For each zone, a value of 1 was assigned for the ring variable to which the zone belonged and a value of 0 was assigned for the other ring variables. These ring variables capture the effect of miscellaneous attractiveness (or unattractiveness) attributes associated with geographic location.

Zonal Spatial Structure Measure

The zonal spatial structure variable is used for the home-based shopping purpose. This variable is used to accommodate the impact of the location pattern of shopping attraction-end zones. Fotheringham (*15*) provided motivation for the inclusion of such a variable. Consider the attraction-end choice of an individual at zone *i* in the two spatial arrangements in Figure 1. All the possible attraction-ends (Zones 1 through 5) are equally distant from zone *i* and are identical in all other respects (i.e., they all are of the same size and attractiveness). The traditional gravity model would then estimate the same trip interchange volumes from zone *i* to each attraction zone. At a disaggregate choice level, the implication is that the probability of attraction-end choice for a trip produced at zone *i* is the same for all the Zones 1 through 5. However, the positioning of the attraction zones relative to one another may have an impact on the choice probabilities and hence on aggregate trip interchanges. One possibility is that the choice probability of Zone 1 may be higher in the first spatial configuration than in the second because of competition effects. Zone 1 may occupy a unique location in the cognitive perception of individuals in zone *i* in the first spatial pattern. Equivalently, Zone 1 competes less with other zones in the first pattern, and there is more competition among other potential zones (2 through 5). Also, Zone 1 may be more attractive in the first configuration because individuals may want to avoid congestion costs in and around a group of zones with several complementary shopping locations in close proximity to each other (*16*). An alternative possibility is that the choice probability of Zone 1 is higher in the second spatial pattern than in the first because of agglomeration effects. The presence of several closely clustered opportunities for shopping may provide individuals at zone *i* with a perception of greater variety (even if this were not actually the case) or more opportunity for comparative shopping. In practice, either competition effects or agglomeration effects may be present and the appropriate effect can be inferred from estimation. To do so, the proximity of a candidate attraction-end zone *j* to other shopping opportunities is specified by using a Hansen-type accessibility measure (*15*):

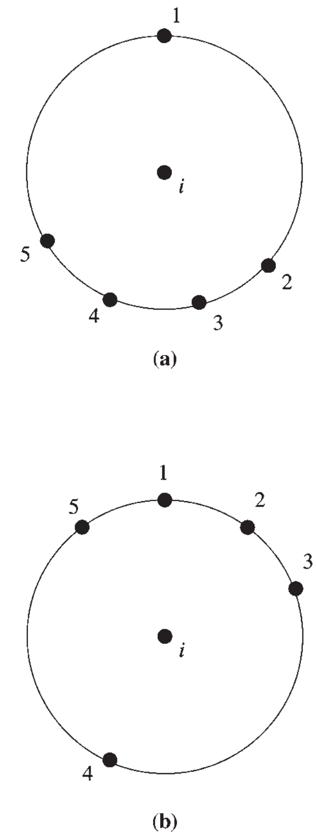

FIGURE 1 Spatial structure effects: (*a*) spatial arrangement 1, (*b*) spatial arrangement 2.

$$M_j = \left[\frac{1}{L} \sum_{l=1}^{L} \left(\frac{\log R_l}{\log H_{lj}} \right) \right] \tag{7}$$

where

R_l = sum of retail and service employment in zone l (a proxy for shopping opportunities in zone l),
H_{lj} = composite travel impedance between zones l and j, and
L = total number of zones in the Boston metropolitan area.

Large values of the proximity variable indicate more opportunities to shop in close proximity of that zone, and small values indicate zones that are spatially isolated from other shopping opportunities. As in the competing destination formulation of Fotheringham (*12*) and Borgers and Timmermans (*17*), the utility of an attraction-end zone j is specified as a linear function of the proximity variable as φM_j, where φ is a spatial structure parameter. If $\varphi < 0$, zones in close proximity to other shopping opportunities have a lower utility than zones in spatial isolation and competition forces dominate. On the other hand, if $\varphi > 0$, zones in close proximity to other shopping opportunities have a higher utility than zones in spatial isolation and agglomeration forces dominate. If $\varphi = 0$, this indicates either absence of spatial structure effects or equally strong agglomeration and competition effects that cancel each other.

Interaction of Sociodemographic Variables with Other Variables

Previous studies have suggested that zonal attributes and impedance measures might interact with sociodemographic characteristics of the individual in determining attraction-end choice. A consistent finding in the geographical literature is that women work closer to home than men because of, among other things, household and childcare responsibilities (*18,19*). Similarly, Madden (*20*) indicated that age appears to have a negative effect on commuting time, possibly because of physiological considerations that increase the sensitivity of older people to traveling long distances. There also is considerable evidence (*21,22*) that higher-income earners travel longer commuting distances. This may be because commuting is a household cost and it makes economic sense for individuals with low income-earning potential to find jobs close to their residences. This discussion suggests that we explore interactions of the composite impedance measure with sex, age, and income of the individual for the home-based work purpose. Although there is less research on the interaction effects of socioeconomic attributes with travel impedance for home-based shopping, these effects were explored in this analysis. Also considered was the interaction of the socioeconomic attributes with zonal attractiveness/location indicators to accommodate any differential sensitivities of different population groups to zonal attributes.

The interaction effects involving age and income were introduced by creating dummy variables characterizing different ranges of age and income and interacting these dummy variables with the composite impedance measure and zonal attractiveness/location indicators. Such a specification does not constrain the age and income effect to be linear or monotonic. Further, the use of dummy variables facilitates the application of the estimated model in forecasting because it is easier to predict trip productions by age and income categories than it is to predict trip productions by the continuous age and income value of individuals.

EMPIRICAL RESULTS

Home-Based Work Attraction-End Choice

The column Home-Based Work Purpose in Table 1 provides the parameter estimates and associated *t*-statistics for the work purpose. The variables that significantly impact work attraction-end choice include the impedance variable, the composite size variable, the dummy variable for Ring 4, and the sociodemographic interactions with the impedance variable.

The parameter signs on all variables are in the expected direction. A larger impedance between the production zone and a candidate attraction-end zone makes it less likely that the candidate attraction-end zone will be chosen. The composite size measure is represented by the total zonal employment; the parameter on total zonal employment is normalized to 1 because it is the only size measure employed in home-based work modeling. The sign on the size variable parameter indicates that zones with high total employment are more likely to be chosen as the attraction-end (relative to zones with low total employment). The coefficient on the size variable represents an inclusive value characterizing the presence of common unobserved zonal attributes affecting the utility of elemental alternatives within a zone. The parameter is close to 0, indicating there are several unobserved zonal attributes that have a common effect on elemental attractions within the zone. The parameter on the Ring 4 dummy variable shows a lower attraction-end choice utility for zones in Ring 4 relative to zones in other rings.

The empirical results indicate significant sociodemographic interactions with the composite impedance variable. (The sociodemographic interactions with Ring 4 and unemployment rate were not statistically significant.)

Home-Based Shopping Attraction-End Choice

The column Home-Based Shopping Purpose in Table 1 presents the estimation results for the home-based shopping purpose.

The sign on the composite impedance measure in Table 1 is as expected. A comparison of the impedance coefficient between the shopping and work purposes indicates that there is much greater sensitivity to travel impedance for shopping relative to work. The coefficient on the composite size variable for the shopping purpose is significantly different from 1, indicating that there are unobserved zonal attributes affecting the utility of elemental destinations within the zone. Among the size variables characterizing the composite size measure, the parameter on zonal area is larger than that on the sum of retail and service employment (the coefficient on retail plus service employment is constrained to 1 for identification). The magnitude on the zonal area variable indicates that 2.6 km^2 (1 mi^2) of zonal area is equivalent to about 5.18 units of retail plus service employment in terms of zonal size representation. The mean value of retail plus service employment in the sample is 1,076 and that of zonal area is 6.8 km^2 (2.62 mi^2). Effectively, then, zonal area contributes substantially less to the composite size measure than does the retail plus service employment measure. The sign on the Ring 4 dummy variable indicates that the utility of a zone not in Ring 4 is greater than a zone in Ring 4. The spatial structure measure is highly significant in its effect on attraction-end choice. The negative parameter on this measure reflects the presence of competition forces; that is, zones in close proximity to other shopping opportunities have a lower utility than zones in spatial isolation.

The sociodemographic interactions with composite impedance are statistically significant (the interactions of crime rate and Ring 4 with sociodemographics turned out to be statistically insignificant). No differences were found in sensitivity among different age groups below 65 years, but individuals older than 65 are more sensitive to impedance than their younger counterparts. Women are more sensitive to impedance than men and individuals in higher-income brackets are less sensitive to impedance than individuals in lower-income brackets.

Evaluation of Fit

The fit of the disaggregate choice specifications were evaluated with the disaggregate equivalent of the conventional gravity model. The disaggregate equivalent of the conventional gravity model includes

TABLE 1 Empirical Estimation Results

Variable	Home-Based Work Purpose		Home-Based Shopping Purpose	
	Parameter	t-stat.	Parameter	t-stat.
Log of composite impedance[a]	-1.3136	-51.04	-1.9366	-86.62
Log of Composite zonal size measure	0.0485	6.80	0.1360	6.67
Total zonal employment[b]	1.0000	-[c]	-	-
Zonal retail plus service employment[d]	-	-	1.0000	-
Zonal land area (in square miles)	-	-	5.18	9.35
Ring 4 dummy variable	-0.7967	-14.49	-0.9395	-14.08
Spatial structure variable	-	-	-4.3618	-48.00
Socio-demographic interactions with impedance				
Female	-0.2410	-30.81	-0.0299	-3.09
55 yrs < age ≤ 65 yrs	-0.2623	-17.96	-	-
age > 65 yrs	-0.3881	-13.29	-0.0723	-5.08
20K < income ≤ 60K	0.1469	6.93	-	-
income > 60K	0.3443	16.10	-	-
60K < income ≤80K	-	-	0.0782	5.69
income > 80K	-	-	0.1478	11.19
Number of Observations	7337		7963	
Log-likelihood at zero	-14277		-15495	
Log-likelihood at convergence	-10613		-6962	

[a]The unit of the composite impedance variable is equivalent highway in-vehicle travel time (in minutes).

[b]The coefficient on this variable is constrained to one in the home-based work model.

[c] All dashes indicate data not applicable.

[d]The coefficient on this variable is constrained to one in the home-based shopping model.

only the composite travel impedance measure and a single size measure, with the coefficient on the size measure constrained to 1 (*23*). Thus, the choice specifications in Table 1 are more general than the disaggregate equivalent of the gravity model.

The fit of the models was examined in both an estimation sample (used in estimation) and a holdout sample (which is not used in estimation). The overall trip samples used earlier were split into an estimation sample (about two-thirds of the trip sample) and a validation sample (one-third of the trip sample). Six additional candidate zonal alternatives were generated for each trip for a total of seven alternatives in the choice set. A measure of fit of a model in the estimation sample is the $\bar{\rho}^2$ value (referred to as the adjusted likelihood ratio index or McFadden's adjusted R^2; see Windmeijer [*24*]) defined as follows:

$$\bar{\rho}^2 = 1 - \frac{L(\hat{\beta}) - Q}{L(0)} \tag{8}$$

where $L(\hat{\beta})$ and $L(\text{O})$ are the log-likelihood function values at convergence and at equal shares, respectively, and Q is the number of parameters estimated in the model. From a formal statistical fit standpoint, the choice specifications in Table 1 are compared with the gravity model by using a nested likelihood ratio test.

In the validation sample, the choice specifications in Table 1 are compared with the gravity model by using the predictive adjusted likelihood ratio index. This measure is computed by calculating the predictive log-likelihood function value at the parameter estimates obtained by maximizing the estimation likelihood function and then computing the corresponding predictive adjusted likelihood ratio index.

The measures of fit in the estimation sample are provided in Table 2. The disaggregate choice specification has a higher adjusted likelihood ratio index compared to the gravity model for both the work and shopping purposes (retail plus service employment is used as the single size measure for the shopping purpose in the gravity model). The nested likelihood ratio index statistic for testing the choice specifications with the gravity model are 2,247.58 and 2,145.80 for the work and shop purposes, respectively. A comparison of these statistics with the chi-squared value with 7 degrees of freedom for the work purpose and 8 degrees of freedom for the shop purpose indicates that one can reject the gravity model at any reasonable level of significance.

Table 3 provides the results of the validation exercise. The predictive adjusted likelihood ratio index is presented in the final row of the table and indicates again that the disaggregate choice model clearly outperforms the disaggregate equivalent of the gravity model. This confirms that the results obtained from the estimation sample are not an artifact of overfitting and are stable.

Another interesting point that emerges from Tables 2 and 3 is that the home-based shopping models perform substantially better than their counterparts for the home-based work purpose.

APPLYING CHOICE MODEL TO OBTAIN TRIP INTERCHANGES

The disaggregate choice model results presented in Table 1 can be applied in a straightforward way to obtain aggregate trip interchanges. To illustrate this procedure, define V_{ijs} as the utility presented by attraction-end zone j for a trip production from zone i made by an individual in sociodemographic group s. Let H_{ij} be the composite travel impedance from zone i to zone j, let D_j be the composite size measure for zone j, and let Z_j be a vector comprising the Ring 4 dummy variable for the work purpose and the Ring 4 as well as the spatial structure measure for the shopping purpose. V_{ij} then can be written as

$$V_{ijs} = \alpha_s \ln H_{ij} + \eta \ln D_j + \mu' Z_j \tag{9}$$

The coefficient on the impedance variable is subscripted by s because the impedance coefficient is a function of sociodemographics. However, the coefficients on the size measure and the Z_j vector are independent of sociodemographics, as obtained in the estimation results. The probability that zone j is selected as the attraction-end for the trip from production zone i made by an individual in a sociodemographic group s can be written as

$$P_{ijs} = \frac{e^{V_{ijs}}}{\sum_k e^{V_{iks}}} = \frac{H_{ij}^{\alpha_s} D_j^{\eta} e^{\mu' Z_j}}{\sum_k H_{ik}^{\alpha_s} D_k^{\eta} e^{\mu' Z_k}} \tag{10}$$

where k is an index of all feasible attraction-end zones for production zone i. This probability is constant across all productions from zone i made by individuals in demographic group s. Thus, if O_{is} is the total number of trip productions from zone i made by individuals in sociodemographic group s, the number of these trips attracted to zone j is given by

$$T_{ijs} = O_{is} P_{ijs} = O_{is} A_{is} H_{ij}^{\alpha_s} D_j^{\eta} e^{\mu' Z_j} \tag{11}$$

where

$$A_{is} = \left[\frac{1}{\sum_k H_{ik}^{\alpha_s} D_k^{\eta} e^{\mu' Z_k}} \right]$$

TABLE 2 Measures of Fit in Estimation Sample

Summary Statistic	Home-Based Work Purpose		Home-Based Shop Purpose	
	Disaggregate choice specification	Disaggregate equivalent of gravity model	Disaggregate choice specification	Disaggregate equivalent of gravity model
Log-likelihood at zero	-9517.54	-9517.54	-10330.94	-10330.94
Log-likelihood at convergence	-7064.33	-8188.12	-4618.32	-5691.22
Number of parameters	8	1	9	1
Number of Observations	4891	4891	5309	5309
Adjusted likelihood ratio index	0.2569	0.1396	0.5521	0.4490
Nested likelihood ratio test statistic	2247.58		2145.80	

TABLE 3 Measures of Fit in Validation Sample

Summary Statistic	Home-Based Work Purpose		Home-Based Shop Purpose	
	Disaggregate choice specification	Disaggregate equivalent of gravity model	Disaggregate choice specification	Disaggregate equivalent of gravity model
Log-likelihood at zero	-4759.75	-4759.75	-5164.50	-5164.50
Log-likelihood at convergence	-2536.40	-2382.00	-2344.92	-2889.44
Number of parameters	8	1	9	1
Number of observations	2446	2446	2654	2654
Predictive adjusted likelihood ratio index	0.2526	0.1382	0.5442	0.4403

The total interchange from zone *i* to zone *j* then can be computed by summing this expression over all sociodemographic groups *s*

$$T_{ij} = \sum_s O_{is} A_{is} H_{ij}^{\alpha_s} D_j^{\eta} e^{\mu' Z_j} \tag{12}$$

SUMMARY AND CONCLUSIONS

The last few decades have seen a shift from the aggregate (or zonal) level paradigm toward the disaggregate (or individual/household level) modeling paradigm within the context of the four-step transportation planning process. Much of this advance in disaggregate modeling has been confined to the trip production and mode choice modeling steps of the planning process. In contrast, the shift toward disaggregate procedures has yet to influence the fundamental specification of trip attraction-end and trip distribution models; in particular, aggregate models continue to be used for these two components of the transportation planning process. This research formulates and estimates (disaggregate) attraction-end choice models that will facilitate the replacement of the aggregate trip attraction models and the aggregate trip distribution models currently used by most urban metropolitan planning organizations.

The research estimated attraction-end models for two trip purposes: home-based work and home-based shopping and personal business. The sample used in the analysis was drawn from the 1991 Boston Household Travel Survey and from supplemental data sources.

The relative performances of the disaggregate attraction-end choice model and the disaggregate equivalent of the gravity model were assessed by using an estimation sample and a validation sample. A consistent result that emerged from the different measures of fit was that the disaggregate attraction-end choice model outperforms the gravity model.

ACKNOWLEDGMENTS

This research was funded by the Massachusetts Highway Department under the guidance of the CTPS of the Boston Metropolitan Organization. The authors thank Karl Quackenbush, Ian Harrington, and others of the CTPS for providing valuable input and feedback during the course of the research.

REFERENCES

1. Guensler, R. Critical Research Needs in Land Use, Transportation and Air Quality. *Transportation Research Circular,* Vol. 389, No. 1, 1993.
2. Domencich, T. A., and D. McFadden. *Urban Travel Demand: A Behavioral Analysis.* North Holland Publishing Company, Amsterdam, 1975.
3. *Models of Mode and Occupancy Choice in the Shirley Highway Corridor.* COMSIS Corp., Silver Spring, Md., 1988.
4. Chang, Y. B., and M. Plager. *Interim Regional Travel Demand Forecasting Model: Development of the 1993 Version.* Memorandum, Central Transportation Planning Staff, Boston Metropolitan Planning Organization, 1994.
5. McClennen, J., K. Quackenbush, and J. Gallagher. *Chronology of FYA Transfer Penalties Research.* Memorandum, Central Transportation Planning Staff, Boston Metropolitan Planning Organization, 1994.
6. Jones, P. M. *Destination Choice and Travel Attributes, Determinants of Travel Choice* (D. A. Hensher and Q. Dalvi, eds.). Saxon House, England, 1978.
7. McFadden, D. *Modeling the Choice of Residential Location: Spatial Interaction Theory and Planning Models* (A. Karlquist, L. Lundquist, F. Snickbars, and J. W. Weibull, eds.). North Holland Publishing Company, Amsterdam, 1978, pp. 75–96.
8. Ben-Akiva, M., G. H. Gunn, and L. Silman. Disaggregate Trip Distribution Models. *Proc., Japanese Society of Civil Engineers,* 1984.
9. Daly, A. Estimating Choice Models Containing Attraction Variables. *Transportation Research,* Vol. 16B, 1982, pp. 5–15.
10. Ben-Akiva, M., and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand.* The MIT Press, Cambridge, Mass., 1985.
11. Govindarajan, A. H. *Disaggregate Attraction-End Choice Modeling Home-Based Work and Home-Based Shopping Trips.* September, 1998.
12. Fotheringham, A. S. (1986). Modeling Hierarchical Destination Choice. *Environment and Planning,* Vol. 18A, pp. 401–418.
13. Department of Employment and Training. *Employment Review.* State of Massachusetts, 1992.
14. *Crime in Massachusetts.* Massachusetts State Police Crime Reporting Unit, 1992.

15. Fotheringham, A. S. Some Theoretical Aspects of Destination Choice and their Relevance To Production-Constrained Gravity Models. *Environment and Planning*, Vol. 15A, 1983, pp. 1121–1132.
16. Lo, L. Substitutability, spatial structure, and spatial interaction. *Geographical Analysis*, Vol. 21, 1991, pp. 132–146.
17. Borgers, A., and H. Timmermans (1987). A Context-Sensitive Model of Spatial Choice Behavior. *Behavioral Modeling in Geography and Planning* (R. Golledge and H. Timmermans, eds.). Croom Helm, London, 1987.
18. Dubin, R. Commuting Patterns and Firm Decentralization. *Land Economics*, Vol. 67, 1991, pp. 15–29.
19. Hodge, D. C. Geography and the Political Economy of Urban Transportation. *Urban Geography*, Vol. 11, 1990, pp. 87–100.
20. Madden, F. J. Why women work closer to home. *Urban Studies*, Vol. 18, 1981, pp. 181–194.
21. Blumen, O. Gender Differences in the Journey to Work. *Urban Geography*, Vol. 15, 1994, pp. 223–245.
22. Hansen, S., and G. Pratt. Geographic Perspective on the Occupational Segregation of Women. *National Geographic Research*, Vol. 6, 1990, pp. 376–378.
23. Bhat, C. *Travel Demand Modeling Improvements: Disaggregate Destination Choice Modeling.* Massachusetts Highway Department, 1995.
24. Windmeijer, F. A. G. Goodness-of-Fit Measures in Binary Choice Models. *Econometric Reviews*, Vol. 14, 1995, pp. 101–116.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.