Trip Generation for Shopping Travel

KWAKU AGYEMANG-DUAH, WILLIAM P. ANDERSON, AND FRED L. HALL

The effect of the geographic location of households on weekday, home-based shopping trips in the greater Toronto area (GTA) is reported. Five zones within the GTA were chosen to reflect different types of location and accessibility. An ordered response model, which maintains the ordinal nature in trip-making decisions, was used in the analysis. The statistical results show that, after controlling for a household's socio-demographic characteristics, a household's location within the metropolitan area has some effect on its weekday, home-based shopping trip generation. In particular, households located in the older urban area are likely to make fewer trips than those living in the suburbs.

The relative importance of discretionary travel (defined as all non-work travel for shopping or social or recreational purposes) has grown over the years and has also captured the attention of both policy makers and transportation demand modelers. In large metropolitan areas, the ratio of discretionary trips to mandatory trips (work and school) is often greater than 1 (*1*). In the Transportation Tomorrow Survey (TTS) in the greater Toronto area (GTA) in 1986, 68 percent of all household trips were for discretionary purposes. The National Personal Transportation Survey in the United States indicated that the number of discretionary trips grew faster than the number of work trips between 1977 and 1988, with discretionary trips making up three-fourths of all household trips in 1988 (*2*). A recent study in the regional municipality of Ottawa-Carleton, Canada, showed that shopping, leisure, and social trips accounted for more than 52 percent of total trips (*3*).

Despite their sheer volume, discretionary trips have been treated crudely in most operational models. For instance, one way to estimate the number of discretionary trips is by applying a constant factor to the number of work and school (mandatory) trips. Discretionary travel, however, may have different temporal and spatial patterns than mandatory travel. Studies on work and school trips focus on maximum peak periods because their purpose is primarily to aid in facility design. The bulk of discretionary trips, however, take place after the morning and evening rush periods when most work trips are over (*4*). Compared with work and school trips, the number of discretionary trips may be more sensitive to such factors as the cost of travel, accessibility, or the land use pattern, all of which tend to vary spatially within a metropolitan area.

In light of the ongoing shift in the focus of transportation planning from plans to build more infrastructure to plans aimed at modifying travel behavior, the development of better models of discretionary travel should be high on the transportation research agenda. The purpose of this paper is to start moving toward improved trip generation models for discretionary travel that are more responsive to locational factors.

The rest of the paper is organized as follows. The next section provides the background for the study. It contains a review of the Urban Transportation Modeling System (UTMS) methods for trip generation analysis and of some past studies of the relationship between trip frequency and the location of trip makers. The next sections discuss the following: (*a*) the data and the rationale for selecting the locations used; (*b*) a brief description of the analytical method used and how it addresses the weaknesses identified in the UTMS approaches; (*c*) a discussion of the statistical results; and (*d*) conclusions.

Department of Geography, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada.

STUDY BACKGROUND

Three things are discussed in this background section: current modeling approaches; the nature of explanatory variables currently used for discretionary trip generation; and recent studies that directly address the relationship between trip generation and location.

Modeling Approaches

Regression models and category analysis are the two main methods used for trip generation in the UTMS. Regression models treat the number of trips generated per household (or individual) as a linear function of a set of explanatory variables. Category analysis divides households into categories on the basis of a cross classification of their characteristics and applies a constant trip generation rate for each category. Both methods have a number of shortcomings.

One problem with the standard regression model is the lack of any built-in upper limit to household trips as the values of explanatory values, such as household size and vehicle ownership, increase. There is also the possibility of the regression models predicting negative trips. In an attempt to deal with these problems, the regression model is sometimes given a probabilistic interpretation. Greene (*5*) has noted, however, that such a model can predict probabilities greater than 1 or less than 0.

The difficulty with category analysis is the lack of any effective way to choose the best groupings of household characteristics and hence the best categories. One way is to minimize the standard deviations among the categories. In situations in which there are many variables and hence many categories, this involves extensive trial and error. Hutchinson (*1*) describes a study by Vandertol using trip data from Hamilton, Ontario, Canada, that produced wide margins of error for households within various categories. The error margins range from 10 percent of the average trip rate for one-worker households to 37 percent for four-worker households. (Although the analysis was based on work trip data, it illustrates the problem of defining the best categories.) Another drawback of category analysis is the lack of inferential statistics. In the absence of such measures, there is no way to assess the statistical significance of the explanatory variables in trip generation.

A problem with both models is that they treat the number of trips per household as a continuous dependent variable. One can of course make a statistical defense of this, but to develop a behavioral basis for trip generation, the dependent variable must be discrete rather than continuous. One possible solution to this problem is to use the poisson regression model in place of the linear regression model. The poisson regression model has been shown to be appropriate in applications to count data, especially when the count for some observations is small or 0 (*6*). An alternative solution is to use one of the family of discrete choice models, which are based on a probabilistic theory of choice among a finite set of options.

Additionally, there is a definite order to the trip-making decision. If a person makes two trips, that person also necessarily makes one trip. The ordinal nature of the trip-making decision is not, and cannot be, captured by either of the regression or category approaches or by the Poisson regression model. The ordered categorical property of the outcomes of the trip-making decision makes it imperative to look for an alternative approach that can exploit the ordering of the information. The ordered response model, a type of discrete choice model that maintains the ordinal nature in the dependent variable in situations in which there are more than two responses, is therefore the best candidate for trip generation analysis. This approach is adopted in this study.

Nature of Explanatory Variables

The types of explanatory variables that are usually used in regression models and category analysis are either the socioeconomic characteristics of households within a zone (for example, income, car ownership, family size), or if these are not available, the characteristics of the zone itself (for instance, population and employment densities). Although cost of travel, accessibility or locational factors have been identified as influencing travel decisions, they are generally excluded from operational models. Ortuzar and Willumsen (*7*) report that attempts to incorporate accessibility measures into UTMS trip generation models have been unsuccessful, noting that the accessibility index is either nonsignificant or has the wrong sign in regression models.

A good indication of the range of explanatory variables currently in use is found in the extensive compilation of trip generation rates by ITE in 1987 (*8*). For generation of shopping trips from residential neighborhoods, for instance, regression models included in the ITE report use household size, the number of vehicles, and the number of dwelling units as explanatory variables. (The report does not distinguish between mandatory and discretionary trips so it is assumed that the model is applicable to all types of trip.) There is a suggestion in the ITE report that location might affect trip generation, but that is not explicitly followed up in the regression result tables. ITE noted that

> dwelling units that were larger in size, more expensive, or farther away from the central business district had a higher trip generation rate per unit than those smaller in size, less expensive, or closer to the CBD. However, other factors, such as geographic location and type of adjacent and nearby development, also had an effect on the trip generation rates. (*8*, p. 256)

The ITE trip generation rates employ adjustment factors for household size, vehicles owned, and density (dwelling units per acre). Although density might be correlated with distance from the central business district, the regression models used to produce the ITE trip generation rates do not take explicit account of location within the city.

Past Studies

Very few studies investigate the relationship between observed trip frequency and location within the city. Two studies that do are reviewed here. The first one is a study carried out in the Canadian regional municipality of Ottawa-Carleton (*3*). The objective of the study was to explore the observed relationship between transportation, land use, and the environment. (The review of the IBI study in this paper concentrates only on the relationship between trip rates and location within the study area.) The study region was divided into nine distinct areas according to similarity in land use mix and density patterns. The major conclusion is that there are no significant differences in the total trip (both work and nonwork) generation rate among the nine areas. The mean daily trips per person range from 2.57 to 3.11 in the nine areas.

This conclusion may be questioned on several grounds. For example, the dispersion of trip rates within areas may vary more than the mean number of trips. Additionally, a different conclusion might have been reached if separate analyses were done for mandatory and discretionary trips because the former is fairly inelastic to locational factors whereas the latter may not be. Thus the results do not really exclude the possibility of some variation in trip-making behavior over space—especially for discretionary trips.

Friedman et al. (*9*) examined trip frequency in older neighborhoods and the newer suburbs in the San Francisco Bay Area. Using 1980 travel data, the study revealed that the number of total trips per household in the two areas differs significantly: 9 and 11 trips for the older neighborhoods and the suburbs, respectively. The study failed to address the following two questions: To what extent does household size correlate with suburban living? Is the difference in trip frequency associated with differences in car ownership in the two areas? (This last question is important because the researchers reported marked differences in mode split for the two areas: 86 percent of trips were by automobile in the suburbs versus 64 percent in the older neighborhoods.) For this reason it is impossible to determine whether the results indicate a "pure" locational effect on trip generation or simply reflect differences in household characteristics over space.

This review has shown that there are problems with the existing approaches to modeling trip generation and that the results of studies on the trip generation-location relationship are inconclusive. The analysis that follows constitutes an attempt to address some of the methodological problems mentioned above and to provide new empirical evidence on the effect of locational factors on discretionary trip making.

DATA

The data for the analysis were obtained from the Transportation Tomorrow Survey conducted between September and December 1986 by the Joint Program in Transportation Studies, University of Toronto, and supported by the Ontario Ministry of Transportation. The TTS was a telephone interview survey of a random sample of 1.5 million households in the GTA, Canada. Completed, usable surveys were obtained for 61,453 households. The GTA is an expanded metropolitan definition that contains 3 of the 25 census metropolitan areas in Canada: Toronto census metropolitan area (CMA) and two contiguous CMAs: Oshawa and Hamilton. For the purpose of the TTS, the GTA was divided into 46 macrozones.

The survey collected data on the sociodemographic characteristics and weekday travel patterns of households. Household characteristics of interest for this analysis are household size, which is the number of persons in the household; the number of household members who are fully employed outside the home, who are employed part time outside the home, who are working at home, and who are unemployed; the number of children (under 16 years); the number of vehicles and the zone of residence. Unfortunately, the survey does not include information on household income or occupation. Census data on income for 1986 for each zone are available in 10 categories. These data cannot be used for detailed analysis, however, because an income level for each household is needed. Data on average zonal income, which are also available, were considered too gross and therefore unsuitable to use in the analysis. The total number of weekday, home-based shopping trips made by automobile and transit is used to calculate household trip generation rates. There were no walk (shopping) trips in the data set for the five zones studied. It was decided to include only home-based shopping trips in the analysis to allow a more direct behavioral interpretation of the results than would be possible with a broader definition of discretionary trips.

The use of observed trip rates raises the question of latent shopping travel demand. This is particularly so when the data used in the analysis were collected for one weekday, despite the fact that many shopping trips take place at the weekend. The unavailability of weekend shopping trip data, however, is less of a problem given that the goal of this study is to search for improved trip generation models rather than to predict the total number of shopping trips.

The use of weekday, home-based shopping trips raises the question of whether these trips constitute a major proportion of total shopping trips. TTS Report 5 provides a table of total (weekdays) shopping trips from each zone. As indicated in Table 1, home-based shopping trips are a relatively constant fraction of the total number of weekday shopping trips. Home-based shopping trips as a percentage of total shopping trips vary from 59 to 66 percent in the five zones, confirming the importance of home-based trips and the need to study them. One should note, however, that shopping trips are defined in the TTS report as all trips that have their destination purpose as shop. It is not clear whether this definition includes trip chains.

The principal hypothesis of this study is that geographic location is an important factor in determining trip generation rates. One simple reason is that location affects people's accessibility, defined as the ease of travel between one point and a set of other points. Ideally, household accessibility measures would have been included in the analysis. However, the data for calculating the accessibility indexes were not readily available. The location of each household in one of five zones within the urban area is therefore used as a proxy for accessibility—although it may also reflect other spatially variant factors such as "lifestyle" differences. Five zones were chosen to reflect different types of location and accessibility (Figure 1).

TABLE 1 Comparison of Weekday Home-Based and Total Shopping Trips per Household

Zone	Home-based	Total	%*
1	0.18	0.30	60
2	0.16	0.27	59
3	0.31	0.47	66
4	0.31	0.49	63
5	0.35	0.57	61

*percentage of weekday home-based shopping trips to total number of shopping trips reported.

Two zones (1 and 2) are within the older urban area and are well served with public transportation, including buses, trolleys, and subways. A third zone is in the inner postwar suburbs, and it is also well served by the transit system. Zones 4 and 5 represent locations that are recently developed suburbs superimposed on older towns. Each of the last three zones has good expressway access. Zones 4 and 5 have, in addition, a network of rural roads but relatively poor public transportation service. The total number of households interviewed in the TTS in the five planning zones were 10,867. Table 2 gives a profile of the five zones.

ORDERED RESPONSE MODEL

The model presented in this section is similar in structure to the probit model developed by McKelvey and Zavoina (*10*) for the analysis of Congressional voting on the 1965 Medicare Bill and by Bhat and Koppelman (*11*) for modeling household income and employment, but with a different set of assumptions. The ordered response model is an extension of the better-known binomial and multinomial logit models. The binomial logit model is used to predict the probability that a categorical variable will take on one of two possible values. In this case it does not matter whether the variable is measured on an ordinal or a nominal scale. The multinomial logit model predicts probabilities for three or more values that a categorical variable can take on. In this case, it is assumed that the variable is measured on a nominal scale. (A common application is the choice among three or more travel modes.) The ordered response model is appropriate when the categorical variable takes on three or more possible values that are subject to some logical ordering. For example, the categorical variable may be successive levels of educational attainment, ratings from an opinion survey, or employment status (unemployed, part-time employed, and full-time employed.) The number of trips generated from a household is clearly such an ordinally scaled categorical variable.

The ordered response model is based on the definition of an abstract score for each household, which can be interpreted in this application as the utility derived by a household from making shopping trips.

$$U_n = V_n + \epsilon_n \tag{1}$$

where

U_n = "total" utility that household n derives from making trips,
V_n = systematic or "observed" utility, and
ϵ_n = random component.

The V_n is defined as a linear function of attributes of the household:

$$V_n = \beta X_n \tag{2}$$

where β and X_n are, respectively, a vector of parameters and a vector of household attributes used as independent variables. (A more general specification would include attributes of the choice alternatives in X; however, no such attributes were employed in this analysis.) The random component is the part of the utility that is unknown to the researcher. It reflects the idiosyncrasies and tastes that vary randomly for each household together with the effect of omitted variables or measurement errors (*12*). The ordered response model assumes "local" instead of "global" utility maximization. Local utility maximization implies a choice situation in which each binary

FIGURE 1 Five macrozones in GTA.

decision consists of whether to accept the current value or "take one more" (*13*). The decision maker stops when the first local optimum is reached. Global utility maximization occurs when all alternatives in the choice set are simultaneously considered. The ordered response model was chosen over the ordered generalized extreme value model of Small (*14*), which maximizes global utility because of its simple mathematical structure, which makes it more convenient for applied analysis.

The model also defines a set of "cut points" associated with each of the possible outcomes. For example, suppose a household can make 0, 1, 2, . . . , J trips, where J is a maximum defined through inspection of the data. Define a cut point λ_1 such that household n will make zero trips if U_n is less than λ_1, or in probabilistic terms

$$P_{n0} = Pr(\beta X_n + \epsilon_n \leq \lambda_1) \tag{3a}$$

where P_{n0} is the probability that household n makes zero trips. The probability that the household makes one trip is now defined as the probability that U_n is greater than λ_1 but less than a second cut point λ_2:

TABLE 2 Profile of Five Zones

Indicators	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
No. of hld*	3210	3723	930	2410	594
Children**	0.46	0.30	0.66	0.79	0.74
Avg. hld size	2.70	2.13	3.22	3.17	3.06
Vehicle/hld	1.11	1.05	1.60	1.76	1.86
Avg. trip***	0.18	0.16	0.31	0.31	0.35

* total number of households interviewed
** average number of household members who were under 16 years
*** average number of home-based shopping trips per weekday
hld = household

$$P_{n1} = Pr(\lambda_1 < \beta X_n + \epsilon_n \leq \lambda_2) \tag{3b}$$

or, more generally,

$$P_{nj} = Pr(\lambda_j < \beta X_n + \epsilon_n \leq \lambda_{j+1}) \quad \text{for } j = 1, \ldots, J - 1 \tag{3c}$$

and

$$P_{nj} = 1 - Pr(\beta X_n + \epsilon_n \leq \lambda_J) \tag{3d}$$

Because it is not possible to observe the values of the random components ϵ_n, the empirical model is derived by making an assumption about their distribution. The random components are assumed logistically distributed:

$$F(\epsilon_n) = 1/[1 + \exp(-\mu\epsilon_n)] \tag{4}$$

where μ is a positive scale parameter that is unobservable; therefore it is assumed that $\mu = 1$. Given these assumptions, an explicit form for Equation 3a can be written:

$$P_{n0} = 1/[1 + \exp(\beta X_n - \lambda_1)] \tag{5a}$$

$$P_{n1} = 1/[1 + \exp(\beta X_n - \lambda_2)] - 1/[1 + \exp(\beta X_n - \lambda_1)] \tag{5b}$$

$$P_{nj} = 1/[1 + \exp(\beta X_n - \lambda_{j+1})] - 1/[1 + \exp(\beta X_n - \lambda_j)] \quad \text{for } j = 2, 3, \ldots, J - 1 \tag{5c}$$

$$P_{nj} = 1 - 1/[1 + \exp(\beta X_n - \lambda_J)] \tag{5d}$$

Estimates of β and $\lambda_1 \ldots \lambda_J$ may be obtained using the maximum likelihood method based on a set of observations (households) making 0, 1, . . . , or J trips for which the attribute data in X_n are available. An application of the ordered response model in travel choice situation was the analysis of trip generation behavior of 774 elderly persons in the Washington, D.C., metropolitan area (*15*).

The ordered response model has the following advantages over the standard regression model of trip generation. First, the property that choice probabilities are necessarily between 0 and 1 means that in prediction mode, the model cannot forecast negative or infinite trips. The second advantage is that the model predicts the whole distribution of the response levels unlike the standard regression approach, which will at best predict the mean of the dependent variable. These advantages of the ordered response model are in addition to what was stated earlier: that the model offers a way to exploit the ordering of information.

STATISTICAL ANALYSIS

The discussion of statistical analysis covers three main areas. First, a brief discussion of the variables used in the estimation of the model is presented. This is followed by a comparative analysis of alternative utility specification functions. Finally, the estimated results are discussed including tests of the estimated parameters and a comparative analysis to assess the overall fit of the model and to demonstrate the extent of zonal variation in trip-making behavior indicated by the model.

Variables Used

The total number of home-based shopping trips over a 24-hr period made by all persons in the household is used in the definition of observed probabilities. ("Trip" as used in the paper is defined as a one-way movement between two places.) If a household is observed to make two trips, the observed probability of making two trips is defined as 1 and the probability of making any other number of trips is defined as 0.

The explanatory variables may be put into two groups: household characteristics and zonal dummy variables. The household characteristics include household size, number of vehicles owned by the household, number of children, and employment status of household members. The household size is expected to be positively correlated with the number of trips because it should influence the level of demand for goods or services, or both. The presence of children in the family may have a dual influence on travel. On the one hand, it may lead to some restrictions on the time available for shopping. Alternatively, it may be regarded as a scale factor leading to increased shopping trips. (The inclusion of household size controls for this scale effect to some extent so that one might expect the number of children to have a negative effect.) Vehicle ownership dramatically improves mobility; hence one might expect more trips in a household with more cars.

The four categories of employment status—full time, part time, working at home, and unemployed—may exert different time budget constraints on shopping trips. Full-time and, to some extent, part-time work is expected to have a negative impact on weekday home-based shopping trips. There is no expectation of the nature of effects of working at home on shopping. Two opposing effects of unemployment may be hypothesized. One effect is that the unemployed person has more time and therefore can make more shopping trips. The other hypothesis is that because a person is unemployed, he or she does not have enough money for shopping.

The four zonal dummy variables were introduced into the ordered response model in both additive and interactive manner. Implicit in the use of additive dummy variables is the assumption that zonal effects are independent of the effect of any household characteristic. It is possible that, for example, household size will have a different impact on trip generation in one zone as opposed to another. To test this hypothesis, the zonal dummies were interacted with household size in the model.

Specification and Comparison of Two Utility Functions

Two utility functions were specified, leading to two types of model. In Model 1, the effects of household size, number of children, and number of vehicles are specified as dummy variables. The utility function in Model 2 is a restricted form of Model 1 in which these same variables were entered in generic form. ("Generic form" means that the explanatory factors are treated as continuous variables. Because of the computational difficulties of including large numbers of dummy variables, the employment variables are entered in generic form for both models.) The two models were estimated in STATA Version 3.0, which uses a Newton-Raphson algorithm. There was some difficulty in estimating Model 1 because of the small number of observations for households of a size greater than six, with five or more children, with more than four vehicles or households making five or more trips. These households were dropped from the data set. The omitted observations constitute only 1.5 percent of the whole data set, leaving 10,701 observations for the estimation of the models.

Using a backward stepwise procedure, all the interactive terms were dropped from both models at a significance level of 0.15, which leads to the conclusion that the dummy variables for the zones have an additive, independent effect on trip generation. The variable working at home was also eliminated from the utility functions as a result of a problem of collinearity with full- and part-time employment. The remaining variables were used to estimate the two models for comparison.

A likelihood ratio test was performed to test the hypothesis that the two models are equal. The test statistic used is $-2(L_2 - L_1)$ which is distributed chi-square. L_1 and L_2 are, respectively, log likelihood values for Model Types 1 and 2. A chi-square value of 32.97 with 10 degrees of freedom was found, which is significant at 0.01, indicating that the two models are unequal. Models 1 and 2 have pseudo R^2 values of 0.0463 and 0.0435, accordingly. (Pseudo R^2 for each model is defined as $1 - L(\beta)/L(c)$, where $L(\beta)$ and $L(c)$ correspond to the log likelihood of a model with all parameters and with only constants, respectively). Model 1 was chosen for further analysis because it had a higher log likelihood value, as evidenced in both the pseudo R^2 and the likelihood ratio test.

Estimated Results

There were two runs of Model 1. The first run had all the household size, number of children, and vehicle dummy variables. (Household size variable has a minimum value of 1 and a maximum of 6 and the number of children and vehicles in each ranges from 0 to 4.) Pairwise significance tests were separately performed for the estimated coefficients of household size and number of children and vehicle dummies. The results showed that the coefficients of all the number of vehicle dummy variables are significantly different from each other. However, household-size dummy variables specific to 4 through 6 and the coefficients for children dummy variables specific to 2 through 4 are not significantly different. Dummy variables specific to household size 4 through 6 and number of children 2 through 4 were therefore constrained to be equal, and the model was run again.

The estimated parameters, together with their standard errors and z-values (used rather than t-values because of the large sample size) for the second run are presented in Table 3. The estimated model is highly significant: a likelihood ratio test of the model against the hypothesis that all the coefficients except the cut points are 0 gives a chi-square value of 556 with 16 degrees of freedom.

As one would expect, the dummy variables for household sizes and number of vehicles are significant. The magnitude of the coefficients of these dummies increases with increasing household size and number of vehicles but at a decreasing rate. The implication is that household sizes and number of vehicles have nonlinear effects on discretionary trip generation.

Two of the three categories of employment status are negatively weighted. Full- and part-time employment is significant, which may be symptomatic of time budget constraints on weekday, home-based shopping trips. The relatively high negative coefficient of full-time employment is indicative of the severe limitations that this variable has on home-based, weekday shopping trips. The effect of unemployment is not statistically significant at 0.1.

The estimated parameters for the two dummy variables for children are negative and are significant. In interpreting the negative coefficients for the children dummies, one should not lose sight of the fact that the data were collected on the weekdays between September and December when children of school age were at school. Child care responsibilities might have had some time budget effects on trip making.

Zonal dummies specific to Zones 3, 4, and 5 are positive and significant, implying that these locations have an effect on trip making relative to the Base Zone 1. The coefficient of the dummy variable for Zone 2 is negative and not statistically significant. (The corresponding value for Zone 1 is 0 by construction.) Pairwise significance tests based on a quadratic approximation to the likelihood function were performed to determine whether the coefficients of these zonal dummy variables are equal. The test results indicate that the differences between the dummy variables for zone pairs 3-4, 3-5, and 4-5 are not significantly different from 0. The test, however, rejects the equality constraint imposed on Zone Pairs 2-3, 2-4, and 2-5. The implication is that Zones 3 through 5 show trip-making propensities distinctly different from those of Zones 1 and 2. There is the possibility that the difference in shopping trip frequency among the zones may be partially because of unobserved income effects. The 1986 average household incomes for Zones 1 and 2 are, respectively, Canadian \$32,000 and \$39,000. On the other hand, each of Zones 3 through 5 has a comparatively higher average household income of approximately Canadian \$45,000 (*16*). However, it is not possible to draw any conclusions about the effect of income on shopping trip frequency in the absence of adequate, reliable data.

Assessment of Prediction Ability

The following exercise is conducted to illustrate the ability of the model to predict aggregate trip-making propensities and also to illustrate the contribution of the zonal dummy variables to the pre-

TABLE 3 Ordered Response Model Estimates

Variable name	Coefficient	Standard error	z-values
Cut point specific to			
trips=1(λ_1)	2.429	0.119	20.412
trips=2(λ_2)	3.873	0.125	30.984
trips=3(λ_3)	5.690	0.160	35.563
trips=4(λ_4)	7.135	0.252	28.310
Household size (HHS) dummy variables specific to:			
HHS=2	0.578	0.108	5.351
HHS=3	0.921	0.163	5.661
HHS=4	1.174	0.236	4.975
Household members:			
fully employed	-0.567	0.700	-8.095
working part-time	-0.234	0.084	-2.796
unemployed	0.085	0.064	1.319*
Children (CHD) dummy variables specific to:			
CHD=1	-0.354	0.091	-3.879
CHD=2	-0.533	0.112	-4.749
Vehicles (VEH) dummy variables specific to:			
VEH=1	0.587	0.960	6.114
VEH=2	0.885	0.108	8.184
VEH=3	1.170	0.143	8.192
VEH=4	1.524	0.214	7.126
Zone (ZN) dummy variables specific to:			
ZN=2	-0.019	0.074	-0.259*
ZN=3	0.457	0.099	4.614
ZN=4	0.446	0.077	5.768
ZN=5	0.562	0.112	5.010

Summary statistics

Number of observations	10701
Chi-square	556.3
Degree of freedom	16
Prob > chi-square	0.0000
Log likelihood (c)	-6033.79
Log likelihood (β)	-5755.64
Pseudo R^2	0.0461

z-values = coefficient / standard error
All variables except those marked by asterisk (*) are significant at 0.01
Trips=0, HHS=1, CHD=0, VEH=0 and ZN=1 were normalised to zero

dictive ability of the model. Define A_{kj} as the aggregate probability that households in Zone k generate j trips, calculated as a relative frequency:

$$A_{kj} = \sum_{n \in Z_k} \frac{P_{nj}}{N_k}$$

where

P_{nj} = probability that household n makes j trips,
Z_k = set of all observations in Zone k, and
N_k = number of observations in Zone k.

A_{kj} is calculated for $k = 1, 2, 3, 4, 5$ and $j = 0, 1, 2, 3, 4$. This calculation is done first on the observed number of trips for households in the data and then on the fitted trip-making probabilities for the same households on the basis of the estimated model. For the purpose of comparison, these observed and predicted probabilities are presented in columns 2 and 3 of Table 4. The results suggest that the model should perform well for the purpose of estimating aggregate trip generation from zones.

To assess the contribution of the zonal dummy variables to the accuracy of prediction, the model was reestimated with the zonal dummies omitted from the specification. Aggregate probabilities calculated on the basis of this model are presented in the fourth column of Table 4. There is some zonal variation in these fitted probabilities, which occurs because of differences in household characteristics in various parts of the metropolitan area. However, these probabilities do not correspond to the observed probabilities nearly as well as those calculated from the original model. This indicates that, even after controlling for spatial variations in household characteristics, there are differences in trip-making behavior at

TABLE 4 Observed and Fitted Aggregate Probabilities

Trips	Observed	Model with zonal dummy variables	Model without zonal dummy variables
Zone 1			
0	0.8677	0.8672	0.8470
1	0.0938	0.0967	0.1100
2	0.0319	0.0300	0.0355
3	0.0054	0.0047	0.0056
4	0.0013	0.0014	0.0017
Zone 2			
0	0.8726	0.8733	0.8529
1	0.1001	0.0924	0.1060
2	0.0246	0.0285	0.0342
3	0.0022	0.0044	0.0054
4	0.0005	0.0014	0.0017
Zone 3			
0	0.7854	0.7838	0.8206
1	0.1427	0.1525	0.1281
2	0.0619	0.0526	0.0424
3	0.0077	0.0085	0.0068
4	0.0022	0.0026	0.0021
Zone 4			
0	0.7886	0.7866	0.8185
1	0.1441	0.1511	0.1297
2	0.0532	0.0515	0.0428
3	0.0106	0.0083	0.0068
4	0.0034	0.0026	0.0021
Zone 5			
0	0.7435	0.7452	0.8003
1	0.1842	0.1771	0.1418
2	0.0534	0.0639	0.0479
3	0.0120	0.0104	0.0077
4	0.0069	0.0033	0.0024

Columns may not add to one due to rounding error.

different locations. These differences may be because of differences in accessibility or other spatially variant factors.

CONCLUSIONS

The objective of this paper was to investigate the effects of location on discretionary household trip generation. Data on weekday, home-based shopping trips, socioeconomic characteristics, and location of households in five widely spaced zones in the GTA were obtained from the TTS. Weekday, home-based shopping trips constitute 59 percent or more of total shopping trips in each zone. An ordered response model was used to analyze the data. Household size, number of vehicles and children, employment status and location of households were included as explanatory variables in the analysis.

The results of the analysis, in terms of the likelihood ratio test of all the explanatory variables, suggest that the estimated model is significant. The z-scores indicate that full- and part-time employment and the dummy variables for household sizes, number of children, number of vehicles, and for Zones 3 through 5 produce significant effects on weekday, home-based shopping travel behavior. The significance of the positive coefficients of dummy variables for Zones 3 through 5 suggest that suburban living is positively correlated with weekday, home-based shopping trips. A comparison of observed and fitted values of aggregate probabilities of making 0,1,2,3, and 4 trips for households in all five zones indicates that the model has good predictive ability and that the inclusion of zonal dummy variables contributes significantly to that ability.

Two implications can be identified from this analysis. First, the ordered response model provides a viable methodology for trip generation. The trip-making decision should no longer be treated as a continuous variable or as a dichotomous response but as a multiple response with a natural order. The other implication is that trip-making behavior appears to be sensitive to location within the metropolitan area, even after controlling for spatial variations in observed household characteristics.

There is a need for further refinements in the application of the ordered response model to discretionary trip generation. The most important is probably the use of accessibility measures in place of spatial dummy variables. Accessibility indexes, which take account of travel costs in time, money, and human effort and of the spatial distribution of opportunities offer transportation planners a more direct way to measure the effect of location on trip-making behavior.

ACKNOWLEDGMENTS

This paper represents the results of the early stages of Kwaku Agyemang-Duah's Ph.D. thesis. Essential assistance in making the data available for this research came from Loy-Sai Cheah at the Joint Program in Transportation, University of Toronto, Canada. The financial support of the Natural Science and Engineering Research Council of Canada is gratefully acknowledged.

REFERENCES

1. Hutchinson, B. G. *Principles of Urban Transport System.* McGraw-Hill, Toronto, Ontario, Canada, 1974.
2. Gordon, P. A., A. Kumar, and H. W. Richardson. Beyond the Journey to Work. *Transportation Research A,* Vol. 22A, No. 6, 1988, pp. 419–426.
3. *The Regional Municipality of Ottawa-Carleton: Environment Review of the Regional Official Plan—Background Paper on Environment, Transportation and Land Use.* IBI Group, Feb. 1993.
4. Prevedouros, P. D., and J. L. Schoffer. Trip Characteristics and Travel Patterns of Suburban Residents. In *Transportation Research Record 1328,* TRB, National Research Council, Washington, D.C., 1991, pp. 49–57.
5. Greene, W. H. *Econometric Analysis,* 2nd ed. Maxwell Macmillan, Toronto, Ontario, Canada, 1993, Chapter 21.
6. Guy, C. M. Recent Advances in Spatial Interaction Modelling: An Application to the Forecasting of Shopping Travel. *Environment and Planning A,* Vol. 19, 1987, pp. 173–186.
7. Ortuzar, J. de D., and L. G. Willumsen. *Modelling Transport.* Wiley, London, 1990, Chapter 4.
8. *Trip Generation Rates,* 4th ed. ITE, Washington, D.C., 1987.
9. Friedman, B., S. P. Gordon, and J. B. Peers. The Effect of Neotraditional Neighborhood Design on Travel Characteristics. Presented at 73rd Annual Meeting of the Transportation Research Board, Washington, D.C., 1994.
10. McKelvey, R. D., and W. Zavoina. A Statistical Model for the Analysis of Ordinal Level Dependent Variables. *Journal of Mathematical Sociology,* Vol. 4, 1975, pp. 103–120.
11. Bhat, C. R., and F. S. Koppelman. An Endogenous Switching Simultaneous Equation System of Employment, Income, and Car Ownership. *Transportation Research A,* Vol. 27A, No. 6, 1993, pp. 447–459.
12. Train, K. *Qualitative Choice Analysis: Theory, Econometrics and Application to Automobile Demand.* MIT Press, Cambridge, Mass., 1986.
13. Ben-Akiva, M., and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Predict Travel Demand.* MIT Press, Cambridge, Mass., 1985.
14. Small, K. A. A Discrete Choice Model for Ordered Alternatives. *Econometrica,* Vol. 55, No. 2, 1987, pp. 409–424.
15. Sheffi, Y. Estimating Choice Probabilities Among Nested Alternatives. *Transportation Research B,* Vol. 13B, 1979, pp. 189–205.
16. *Census of Canada 1986: Population and Dwelling Characteristics Profile,* Series 94-112. Statistics Canada, 1986.

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