

The Spatial Analysis of Activity Stop Generation

Chandra Bhat and Huimin Zhao

Department of Civil Engineering - ECJ 6.810

University of Texas at Austin, Austin, Texas, 78712

Phone: (512)232-6272, Fax: (512) 475-8744

E-Mail: bhat@mail.utexas.edu, hzhao@mail.utexas.edu

Abstract

Travel demand analysis is intrinsically spatial; yet spatial analysis considerations are seldom recognized and accommodated in travel modeling. The objective of this paper is to identify the spatial issues that need to be recognized in demand modeling, and to propose a multi-level, mixed logit, formulation to address these spatial issues in the context of activity stop generation. The multi-level model is estimated using the maximum simulated likelihood method. Empirical results obtained from applying the model to study shopping activity stop generation in the Boston metropolitan area are presented and discussed.

Keywords: Spatial analysis, Multi-level modeling, Activity-based analysis, Simulated maximum likelihood, Halton sequences, Accessibility measure.

1. Introduction

Travel demand analysis usually entails the partitioning of the planning area into several zones, each of which represents an aggregate spatial unit from where trips are produced and to where trips are attracted. The spatial and temporal attributes of trip interchanges between two zones are typically estimated using activity-based or trip-based models at an individual (or household) level, and suitable sample enumeration techniques are applied to forecast traffic flows. Three inter-related spatial-analytic issues arise in the estimation of such individual (or household) level models: spatial dependency, spatial heterogeneity, and spatial heteroscedasticity.

Spatial dependency (also referred as spatial autocorrelation) occurs among activity/travel choices of decision-making agents clustered within a zone because of unobserved locational effects (the decision-making agent may be a household or individual). For example, Bhat's (2000a) analysis suggests the presence of unobserved factors influencing work mode choice among individuals living in a particular zone as well as among individuals working in a particular zone. Such unobserved factors might include zone-related attributes such as pedestrian or transit friendliness of a zone, natural or acquired attitudinal factors pervading the decision-making of individuals living close together, or common unobserved demographic attributes of individuals living close together (see Jones and Bullen, 1994; Miller 1999). In general, ignoring spatial dependency can result in mis-estimated standard errors in linear models (Anselin and Griffith, 1988) and (in addition) inconsistent parameter estimation in non-linear models (Case, 1992).

Spatial heterogeneity refers to differences in relationships between the dependent variable of interest and the independent variables across spatial units (or zones) in a study region. Such variations in the relationship are likely because of location-specific effects, as demonstrated by Fotheringham *et al.* (1996, 1997) using visualization techniques. Fotheringham and Brunson (1999)

discuss the reasons for these variations in detail, identifying two equally plausible but indistinguishable sources in analysis. One source is intrinsic behavioral differences in the decision-making of agents across spatial units. The other source is the lack of information (on the part of the analyst) regarding some agent-related or spatial-unit related attributes that affect decision-making. In either case, it behooves the travel analyst to accommodate local variations (*i.e.*, recognize spatial non-stationarity) in the relationship across a study region rather than settle for a single global relationship. In particular, a single global relationship in a study region may not reflect decision-making behavior appropriately in any local part of the study region. Further, this potential mismatch in the global relationship and local relationships can lead to inconsistent estimates of the effect of variables at the global-level if the relationship between the dependent variable of interest and the independent variables is characterized by a non-linear function.

Spatial heteroscedasticity refers to heterogeneity in the variance of the unobserved component across spatial units. For example, consider a variable that captures a measure of accessibility to pursue activities in a zone and that is not included in the model due to data unavailability. It is possible that there is large variation in stop-making across households residing in a zone with very good accessibility (because stop-making in the zone is not very dependent on accessibility and is dictated by other unobserved household characteristics), while there is small variation in stop-making across households residing in a zone with poor accessibility (because of the more uniform tendency to make fewer stops due to the dominance of the effect of low accessibility on stop-making). Ignoring spatial heteroscedasticity when it is present leads to inconsistent parameter estimates in non-linear models (see McMillen, 1992, 1995).

There has been considerable discussion of the three spatial issues of dependency, heterogeneity and heteroscedasticity in the geography literature, as should be evident from the

references provided above. However, there has been little discussion and recognition of spatial analytic issues in the travel demand modeling literature. This paper highlights the need to accommodate spatial issues in travel modeling, and proposes a specific spatial model formulation in the context of activity stop generation.

In the Sections 1.1 and 1.2, we position the current study in the context of the earlier spatial analysis literature, both from a methodological perspective as well as an application perspective.

1.1. Methodological focus of paper

Previous studies in spatial analysis have, for the most part, focused on linear models with a continuous dependent variable. However, the dependent variable of interest in many spatial contexts is of a discrete or limited-dependent form (for example, whether or not an individual adopts a new technology may be influenced by his neighbors or the type of car purchased by a household may be influenced by other households in the neighborhood). Only more recently has there been increasing focus on non-linear models with discrete and limited-dependent variables (see McMillen, 1995; Dubin, 1995; Le Sage, 2000).

The methodological focus of this paper is to contribute to the growing body of literature on accommodating spatial issues within a discrete choice formulation using a multi-level analysis framework. Multi-level analysis (also labeled as variance-components analysis or hierarchical modeling) has been applied in several fields, including education (Goldstein *et al.*, 1993; Langford and Lewis, 1998), aspatial travel modeling (Bhat, 1999; Goulias, 2000), sociology (Hox and Kreft, 1994), health and medicine (Ribaudo *et al.*, 1999), survey analysis (O’Muircheartaigh and Campanelli, 1999), political science (Yang *et al.*, 2000), and geography (Jones and Duncan, 1996; Verheij, 1997; Smit, 1997). The application of multi-level analysis has been particularly burgeoning

in the field of geography since geographical analysis is intrinsically spatial and involves the grouping of elementary units of interest (for example, households and individuals) into higher level spatial clusters (such as neighborhoods, communities, and traffic analysis zones). In such a spatial clustering context, it is important to recognize and differentiate heterogeneity among aggregate spatial clusters (which causes auto-correlation among elementary units within each spatial cluster) and heterogeneity among elementary units. As indicated by Jones and Duncan (1996) and several other researchers, ignoring this differentiation and modeling the behavior of interest at a single level invites the pitfalls of either the ecological fallacy when the level of analysis is solely at the aggregate spatial level (*i.e.*, failing to recognize that it is the elementary units which act and not aggregate spatial units) or the atomistic fallacy when the analysis is pursued entirely at the elementary unit level (*i.e.*, missing the spatial context in which elementary units behave).

Within the context of a multi-level analysis framework, we use a mixed-logit formulation for the model structure. The mixed logit formulation is a very flexible structure and has seen increasing use in aspatial analysis (see Bhat, 1998, 2000b; Revelt and Train, 1998; Brownstone and Train, 1999). In contrast, the formulation has seen little to no use in the spatial choice literature. To the author's knowledge, the only previous application of the mixed logit structure in a spatial context is the study by Bhat (2000a). This current study accommodates more general patterns of spatial dependency and spatial heterogeneity compared to, and adds spatial heterogeneity to, the structure used by Bhat (2000a).

1.2. Application focus of paper

The application focus of this paper is the area of urban travel demand analysis. The field of urban travel demand analysis has seen remarkable progress over the past two decades, but spatial

considerations have seldom been recognized and accommodated in the field. While GIS platforms are increasingly being used by many metropolitan and state agencies for travel analysis, they are being primarily used for data assembly and presentation of model results. The underlying analytic methods have remained aspatial.

Within the broad application context of travel demand analysis, the specific emphasis in this paper is on modeling households' decision on number of out-of-home activity episodes (*i.e.*, stops). Household stop generation is an important component of an overall activity-based analysis framework that emphasizes travel as being derived from the need to participate in activities (for a recent comprehensive review of the activity-based approach see Bhat and Koppelman, 1999). Several earlier studies have focused on household activity stop generation, either in isolation or jointly with other dimensions of choice such as stop allocation to household members, organization of stops within the overall daily activity-travel schedule, and stop attributes such as location, duration and travel time to stop from earlier episode (for recent examples, see Wen and Koppelman, 2000; Misra and Bhat, 2000; Bhat and Singh, 2000; Kitamura *et al.*, 2000). There have also been other studies in the traditional trip-based strand of literature which focus on the related issue of trip production modeling (recent studies in this area include Agyemang-Duah and Hall, 1997; Strambi and van de Bilt, 1998; Washington and Wolf, 1997). The problem with focusing on trips rather than stops is that the dependence among trip attributes and the organization of trips is difficult to accommodate.

The studies of stop or trip generation identified above, and other earlier studies in this area, have not recognized the spatial issues of dependency, heterogeneity, and heteroscedasticity. In this paper, we propose a mixed-ordered logit model for household shopping stops that accommodates

these spatial issues. The proposed model is estimated using a data set from the Dallas-Forth Worth area.

The rest of this paper is structured as follows. Section 2 discusses the model structure. Section 3 presents the model estimation technique. Section 4 describes the data source and empirical results. Section 5 summarizes important findings from the study.

2. Model Structure

The model structure in the current paper takes an ordered-response formulation that recognizes the ordinal nature of stop-making. The ordered-response formulation was initially proposed by McKelvey and Zavoina (1975), and has been used recently by Agyemang-Duah and Hall (1997), Bhat *et al.* (1999), and Bhat (1999) to model number of stops.

In the usual notation of the ordered response structure, let the latent stop-making propensity s_{qz}^* of household q ($q = 1, 2, \dots, Q$) in zone z ($z = 1, 2, \dots, Z$) to make shopping stops be written as:

$$s_{qz}^* = \eta_z + \gamma'_z x_{qz} + \varepsilon_{qz}, \quad s_{qz} = k \text{ if } \delta_{k-1} < s_{qz}^* \leq \delta_k, \quad (1)$$

where k is an index for number of stops ($k = 0, 1, 2, \dots, K$), η_z is a zone-specific scalar constant term for households in zone z , x_{qz} is a column vector of exogenous variables for household q in zone z (x_{qz} does not include a constant), γ_z is a corresponding zone-specific column vector of coefficients, and ε_{qz} is a standard logistic random term that captures the idiosyncratic effect of all omitted variables which are not zone-specific. s_{qz} is the observed number of daily shopping stops made by household q residing in zone z . It is characterized by the stop-making propensity s_{qz}^* and the threshold bounds (the δ 's) in the usual ordered-response fashion.

Equation (1) represents the micro-level model for households. We now allow the zone-specific scalar term η_z and the zone-specific coefficient vector γ_z to vary across zones in a higher-level macro-model.

The macro-level model for the zone-specific scalar term is specified as the sum of an unobserved component α_z and a linear function of observed zone-specific variables y_z : $\eta_z = \alpha_z + \lambda'y_z$, where λ is a column vector fixed across all zones. The variance in α_z captures intercept (or intrinsic) unobserved heterogeneity in household stop-making propensity across zones; that is, it allows differences across zones in the propensity of households to make shopping stops. Since the zones used in estimation are a sample drawn from the overall population of zones in the study region, it is appropriate to consider the α_z terms to be random variables. We consider the α_z terms to be normally distributed, independent of each other, and independent of the idiosyncratic elements of the household captured in ε_{qz} . The mean of α_z is zero (the restriction on the mean is an innocuous one because of the inclusion of the thresholds). The variance of the distribution of α_z may be assumed to be fixed across zones (homoscedastic) or can be permitted to differ across zones (heteroscedastic). The latter formulation is a generalization of the former and may be more appropriate because of the spatial heteroscedasticity considerations discussed in the first section. In this paper, we allow for such differences in intercept unobserved heterogeneity by specifying the variance of α_z to be a function of zonal attributes. That is, $\alpha_z \sim N[0, \sigma_z^2]$ with $\sigma_z = \exp(\omega + \mu'w_z)$, where w_z is a vector of zonal attributes. The exponential functional form is used in the specification to ensure the non-negativity of the standard error (Greene 1997, p. 889; McMillen, 1995; Swait and Adamowicz, 1996) also use an exponential form for accommodating heteroscedasticity in discrete choice models).

The macro-level model for the zone-specific coefficient vector γ_z is specified by writing the j th element of the vector as: $\gamma_{zj} = \gamma_j + v_{zj}$, where γ_j represents the average areawide slope parameter of the effect of the j th variable ($j = 1, 2, \dots, J$) in the x_{qz} vector on household stop-making propensity, and v_{zj} is a zone-specific term representing the deviation of the slope effect from the area-wide slope for zone z . We consider the v_{zj} terms to be normally distributed, independent and identically distributed across zones, and independently distributed across elements j . Thus, $v_{zj} \sim N(0, \Omega_j^2)$ and, therefore, $\gamma_{zj} \sim N(\gamma_j, \Omega_j^2)$.

The reader will note that spatial dependency among households within the same zone is generated by the random terms v_{zj} ($j = 1, 2, \dots, J$) and α_z in the zone-level macro-models. Of course, these terms were motivated in the discussion above from spatial heterogeneity considerations. Also, the heteroscedastic specification for the distribution of α_z generates spatial heteroscedasticity.

The micro-level model of equation (1) may be combined with the macro-level models to form the following multi-level spatial equation:

$$s_{qz}^* = \lambda'y_z + \sum_j \gamma_{zj} x_{qzj} + \alpha_z + \varepsilon_{qz}, \quad s_{qz} = k \text{ if } \delta_{k-1} < s_{qz}^* \leq \delta_k, \quad (2)$$

Conditional on the α_z and γ_{zj} terms ($j = 1, 2, \dots, J$), we get the familiar ordered-response logit form for the choice probability of household q in zone z making k number of stops (L represents the logistic distribution function below)

$$P_{qzk} | (\alpha_z, \gamma_{z1}, \gamma_{z2}, \dots, \gamma_{zJ}) = L(\delta_k - \alpha_z - \lambda'y_z - \sum_j \gamma_{zj} x_{qzj}) - L(\delta_{k-1} - \alpha_z - \lambda'y_z - \sum_j \gamma_{zj} x_{qzj}) \quad (3)$$

The unconditional probability of choosing k number of stops for a household q in a randomly selected zone with observed vectors y_z , x_{qz} and w_z can now be obtained by integrating the conditional choice probabilities in equation (3) with respect to the assumed random (and independent) normal distributions for the $(J+1)$ random variables $\alpha_z, \gamma_{z1}, \gamma_{z2}, \dots, \gamma_{zJ}$. The resulting expression has the following form:

$$P_{qzk} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \left[L(\delta_k - \alpha_z - \lambda' y_z - \sum_j \gamma_{zj} x_{qzj}) - L(\delta_{k-1} - \alpha_z - \lambda' y_z - \sum_j \gamma_{zj} x_{qzj}) \right] dF(\alpha_z) dF(\gamma_{z1}) dF(\gamma_{z2}) \dots dF(\gamma_{zJ}). \quad (4)$$

3. Model Estimation

The parameters to be estimated in the multi-level model of equation (2) include the vector $\zeta = (\omega, \mu, \lambda)'$ and the vector $\xi_j = (\gamma_j, \Omega_j)$ for all $j = 1, 2, \dots, J$. Let $\tau = (\zeta', \xi_1', \xi_2', \dots, \xi_J)'$ represent the full set of parameters to be estimated and let D_z be the set of all households residing in zone z . To develop the likelihood function, we need the likelihood function of the sequence of observed number of stops for all households residing in zone z . Conditional on $\alpha_z, \gamma_{z1}, \dots, \gamma_{zJ}$, this function may be written as:

$$L_z | (\alpha_z, \gamma_{z1}, \dots, \gamma_{zJ}) = \prod_{q \in D_z} \prod_{k=1}^K [P_{qzk} | (\alpha_z, \gamma_{z1}, \dots, \gamma_{zJ})]^{M_{qzk}}, \text{ where} \quad (5)$$

$$M_{qzk} = \begin{cases} 1 & \text{if the } q\text{th household in zone } z \text{ makes } k \text{ stops} \\ 0 & \text{otherwise.} \end{cases}$$

The unconditional likelihood function of the choice sequence is:

$$L_z(\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \{L_z | (\alpha_z, \gamma_{z1}, \dots, \gamma_{zJ})\} dF(\alpha_z) dF(\gamma_{z1}) \dots dF(\gamma_{zJ}). \quad (6)$$

Now define a_z and u_{zj} ($z=1,\dots,Z$, $j=1,\dots,J$) as standard-normal variates so that $a_z = \alpha_z / [\exp(\omega + \mu'w_z)]$ and $u_{zj} = \gamma_{zj} / \Omega_j$. Then, using equations (3) and (5), the unconditional likelihood function of equation (6) may be written for a given value of the parameter vector τ as:

$$L_z(\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \left[\prod_{q \in D_z} \left\{ \prod_{k=1}^K \left[L(\delta_k - [\exp(\omega + \mu'w_z)]a_z - \lambda'y_z - \sum_j \Omega_j u_{zj} x_{qzj}) - \right. \right. \right. \\ \left. \left. \left. L(\delta_{k-1} - [\exp(\omega + \mu'w_z)]a_z - \lambda'y_z - \sum_j \Omega_j u_{zj} x_{qzj}) \right]^{M_{qzk}} \right\} \right] \times \\ d\Phi(a_z) d\Phi(u_{z1}) d\Phi(u_{z2}) \dots d\Phi(u_{zJ}). \quad (7)$$

where $\Phi(\cdot)$ represents the standard normal distribution function. The log-likelihood function is $L(\tau) = \sum_z \ln L_z(\tau)$.

The log-likelihood function involves the evaluation of a $(J+1)$ -dimensional integral (J is the number of variables with random response coefficients). Conventional quadrature techniques cannot compute the integrals with sufficient precision and speed for estimation via maximum likelihood when the dimensionality of the integration is greater than 2.

In the current study, we apply simulation techniques to approximate the integrals in equation (7) and maximize the resulting simulated log-likelihood function. The simulation technique computes the integrand in equation (7) at several values of a_z and u_{zj} drawn from the normal distribution. Specifically, we draw a particular realization of a_z and u_{zj} ($j = 1, 2, \dots, J$) by generating a vector of $(J+1)$ standard normal numbers for each zone z and subsequently compute the integrand in equation (7) for a given value of the parameter vector τ . We then repeat this process N times for each zone for the given value of the parameter vector τ . Let $\tilde{L}_z^n(\tau)$ be the realization of the

zonal likelihood function in the n th draw ($n = 1, 2, \dots, N$). The zonal likelihood function is then approximated by averaging over the $\tilde{L}_z^n(\tau)$ values:

$$\tilde{L}_z(\tau) = \frac{1}{N} \sum_{n=1}^N \tilde{L}_z^n(\tau), \quad (8)$$

where $\tilde{L}_z(\tau)$ is the simulated likelihood function for the sequence of choices of households in zone z given the parameter vector τ . $\tilde{L}_z(\tau)$ is an unbiased estimator of the actual likelihood function $L_z(\tau)$. Its variance decreases as N increases. It also has the appealing properties of being smooth (*i.e.*, twice differentiable) and being strictly positive for any realization of the finite N draws.

The simulated log-likelihood function is constructed as:

$$SL(\tau) = \sum_{z=1}^Z \log[\tilde{L}_z(\tau)] \quad (9)$$

The parameter vector τ is estimated as the vector value that maximizes the above simulated function. Under rather weak regularity conditions, the maximum simulated log-likelihood (MSL) estimator is consistent, asymptotically efficient, and asymptotically normal (see Hajivassiliou and Ruud, 1994; Lee, 1992).

An important issue in the above simulation-based estimation is the approach used to draw realizations for a_z and u_{zj} from their population normal distributions. The typical approach in econometrics uses a random Monte Carlo method in which the realizations are randomly generated from the underlying population distributions. The basic principle in the random method is to replace a continuous average by a discrete average over randomly chosen points. By the strong law of large numbers, convergence is almost sure in this method. Of course, in actual implementation, truly random sequences are not available; instead, deterministic pseudo-random sequences which appear random when subjected to simple statistical tests are used (see Niederreiter, 1995 for a discussion of

pseudo-random sequence generation). This pseudo-Monte Carlo (or PMC) method has a slow asymptotic convergence rate with the expected integration error of the order of $N^{-0.5}$ (N being the number of pseudo-random points drawn from the s -dimensional integration space). Thus, to obtain an added decimal digit of accuracy, the number of draws needs to be increased 100-fold.

More recently, Bhat (2001) proposed the use of a quasi-Monte Carlo (QMC) method to generating the realizations for simulation-based econometric estimation. Rather than using pseudo-random sequences for the discrete points, the QMC approach uses “cleverly” crafted non-random and more uniformly distributed sequences within the domain of integration. The underlying idea of the method is that it is really inconsequential whether the discrete points are truly random; of primary importance is the even distribution (or maximal spread) of the points in the integration space. The sequences used in the QMC method are labeled as quasi-random sequences, though this is a misnomer since randomness plays no part in the construction of the sequences. The convergence rate for quasi-random sequences is, in general, faster than for pseudo-random sequences. In particular, the theoretical upper bound for the integration error in the QMC method is of the order of N^{-1} , where N is the number of quasi-random integration points.

Bhat has studied the empirical performance of the PMC and QMC simulation approaches in the context of a mixed logit model. He used the Halton sequence in the QMC simulation because of its conceptual simplicity. The results from this study indicate that the QMC method out-performs the pseudo-Monte Carlo (PMC) methods for mixed logit model estimation, and provides substantial time savings in estimation. Subsequent studies by Train (1999) and Hensher, (1999) confirm the substantial reduction in computational time for mixed logit estimation using the QMC method. In the current paper, we use the Halton sequence for simulation estimation. Details of the Halton sequence

and the procedure to generate this sequence is available in Bhat, (2001) and the many references cited therein.

The QMC simulation technique to evaluate the integral in the log-likelihood function of equation (9) involves generating a $(J+1)$ -dimensional Halton sequence for a specified number of "draws" N for each zone. To avoid correlation in simulation errors across zones, separate independent draws of N Halton numbers in $(J+1)$ -dimensions are taken for each zone. This is achieved by generating a Halton "matrix" Y of size $G \times (J+1)$, where $G = N*Z+10$ (Z is the total number of zones in the sample). The first ten terms in each dimension are then discarded because the integrand may be sensitive to the starting point of the Halton sequence. This leaves a $(N*Z) \times (J+1)$ Halton matrix which is partitioned into Z sub-matrices of size $N \times (J+1)$, each sub-matrix representing the N Halton draws in $(J+1)$ dimensions for each zone (thus, the first N rows of the Halton matrix Y are assigned to the first zone, the second N rows to the second zone, and so on). In the current paper, we use 150 Halton draws for each zone ($N = 150$), since earlier studies have shown that simulation error even with fewer Halton repetitions is rather small.

All estimations and computations were carried out using the GAUSS programming language on a personal computer. Gradients of the simulated log-likelihood function with respect to the parameters were coded.

4. Empirical Application

4.1. Data source and sample description

The data source used in the current paper is drawn from a household activity survey conducted in 1991 by the Central Transportation Planning Staff (CTPS) in the Boston Metropolitan region. The survey collected data on socio-demographic characteristics of the household and each

individual in the household. The survey also included a one-day (mid-week working day) activity diary to be filled out by all members of the household above five years of age. This activity diary data was used to determine the number of shopping stops generated by each individual in the household, and then aggregated across individuals in the same household to obtain household level stops.

The Boston metropolitan planning area comprises 986 traffic analysis zones. Zone-to-zone travel impedances by travel mode have been generated by the Central Transportation Planning Staff of the Boston Metropolitan Planning Organization. The land-use attributes of each zone have also been determined. These zone-to-zone impedance values and zonal land-use attributes are used in this study to develop a measure of accessibility of each zone to shopping opportunities. Each household's residence is geo-coded to the traffic analysis zone in which the household lies.

The sample for the analysis includes 1815 households whose residence locations are distributed across 495 traffic analysis zones. The sample was confined to zones in which at least two households resided in the original data. Of the 495 traffic analysis zones represented in the sample, 190 zones had two household observations, 126 had three households, 78 had four households, 33 had five households, and the remainder had six or more households.

The distribution of number of shopping stops among the households in the sample is as follows: 0 stops (45.6%), 1 stop (31.9%), 2 stops (14.1%), 3 stops (5.3%), 4 stops (1.8%), and 5 stops (1.3%). These figures indicate a rather high level of participation in shopping, with 54.4% of all households undertaking at least one shopping stop in the day. The distribution of stops in the sample used in the analysis is almost exactly the same as in the original raw data.

4.2. Measure of accessibility

The accessibility of a residential zone j to shopping opportunities is specified using a Hansen-type accessibility measure (Fotheringham, 1993):

$$M_j = \left[\frac{1}{L} \sum_{l=1}^L \left(\frac{\log R_l}{\log H_{lj}} \right) \right], \quad (10)$$

where R_l is the retail plus service employment in zone l (a proxy for shopping opportunities in zone l), H_{lj} is the composite travel impedance between zones l and j , and L is the total number of zones in the Boston Metropolitan area. Large values of the accessibility variable indicate more opportunities to shop in close proximity of that zone, while small values indicate zones which are spatially isolated from shopping opportunities.

The composite travel impedance H_{lj} accommodates multiple impedance measures (in-vehicle travel time, out-of-vehicle travel time, and cost) for each travel mode and accounts for differential modal availabilities for different zone pairs. The formulation for this composite impedance measure is similar to the one used by Bhat, *et al.* (1998). It first entails converting the multiple impedance measures for each mode into a single modal impedance measure. To achieve this, we translate out-of-vehicle time (OVTT) and travel cost (COST) into equivalent in-vehicle time (IVTT) units using estimated coefficients on these variables from mode choice modeling results (see Bhat *et al.*, 1998 for details). The next step in developing the composite impedance measure involves combining the impedance measure for each available mode into a composite travel impedance measure H (for each zonal pair). The universal set of modes in the Boston region include highway, transit, and walk (the bike mode accounts for very few trips for shopping and so is not considered here). The highway mode is available for all zone pairs in the Boston Metropolitan area.

However, the transit and walk modes are not universally available. Our composite impedance formulation uses a parallel conductance formula to combine the mode-specific impedances. To develop this formulation, we define a transit availability dummy variable θ_t (for each zone pair) that takes a value of 1 if transit is available and zero otherwise, and a walk availability dummy variable θ_w that takes a value of 1 if walk is available and zero otherwise. The composite impedance H is then written as:

$$H = (1 - \theta_t)(1 - \theta_w)C + \theta_t(1 - \theta_w) \left(\frac{C}{1 + \frac{C}{T^\lambda}} \right) + \theta_w(1 - \theta_t) \left(\frac{C}{1 + \frac{C}{W^\mu}} \right) + \theta_t\theta_w \left(\frac{C}{1 + \frac{C}{T^\lambda} + \frac{C}{W^\mu}} \right) \quad (11)$$

where C is the highway impedance, T is the transit impedance, and W is the walk impedance. If both transit and walk are unavailable, the first term applies and the composite impedance is just 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. The reader will note that the formulation in equation (11) satisfies the following fundamental principle: if two zones j and m have the same highway travel impedance from a residential zone t , but zone j is served by only a subset of the modes that serve zone m from zone t , then the overall impedance assigned for travel from zone t to zone j should be lower than that assigned for travel from zone t to zone m . The estimation of the λ and μ parameters ($\lambda = 1.6155$, $\mu = 0.9988$) is achieved in a destination choice model. The estimation procedure and the intuitive interpretations of the coefficients are discussed in Bhat *et al.*, (1998).

4.3. Variable specification

Three sets of variables were considered to explain household shopping stop making propensity. These are (a) household employment-related variables, (b) household structure variables, and (c) location and accessibility variables.

The household employment-related variables include the number of full-time, part-time and unemployed individuals, and household income.

The household structure variables identified households in one of the following categories: cohabitation of two adults of opposite sex, couple families, single-person households, single parent families, nuclear families, and other household types. In our analysis, we use nuclear families and other household types as the base category, and introduce dummy variables specific to cohabitation, couple, single-person, and single-parent families. For ease in presentation, we will refer to the base household type category simply as “nuclear family”, because of the dominance of such households in the base category. Additional household structure variables considered in our specifications were presence and number of children in different age groups

The household location variables classified household residence into urban, suburban, and rural locations. These aggregate location indicators, however, did not significantly impact stop-making propensity after inclusion of the more disaggregate accessibility measure for each household based on its zone of residence. But the aggregate location indicators contribute to spatial heteroscedasticity, as we discuss later. Table 1 presents the descriptive sample statistics for the exogenous variables used to explain household shopping stop-making propensity.

4.4. Empirical results

The parameter estimation results for the (aspatial) ordered response logit (ORL) model and the (spatial) mixed ordered response logit (MORL) model are presented in Table 2. The parameter estimates have the same sign in both the models. The effect of the employment-related variables indicate a higher stop-making propensity as the number of adults in the household increases (the sum of number of full-time employed adults, part-time employed adults and unemployed adults is equal to the number of adults). The relative magnitudes of the parameters imply that unemployed adults have a higher shopping stop propensity than adults employed part-time, and adults employed part-time have a higher stop-making propensity than adults employed full-time. This effect reflects the positive impact of time availability on shopping stop-making propensity. The positive influence of income on shopping stop-making propensity may be the result of the increased consumption potential of high income households.

The effect of the household structure variables suggests important differences in shopping stop-making propensity among household types. The ORL model indicates that the presence of children between the ages of 12 and 16 years increases the stop-making propensity of households. The MORL model, which allows the coefficients to vary across zones, indicates substantial variation in the effect of this variable across zones. The estimated mean coefficient on the variable is 0.120 and is statistically insignificant; however, the estimated variation around this mean value is 1.15 and statistically significant. A similar result is obtained for the effect of couple and cohabitation families. The ORL model suggests that couples and cohabitation families have a fixed higher propensity to make shopping stops compared to nuclear family households, while the MORL model indicates substantial variation in this effect across zones and suggests that couples/cohabitation families may have a higher stop-making propensity or a lower stop-making propensity depending on their zone of

residence. Such differences may be the result of micro-design urban-form features within a zone that interact with the family structure to either intensify or suppress stop-making propensity. As indicated recently by Badoe and Miller (2000), interactions between household socio-economics and urban form can play an important role in understanding individual and household activity/travel decision making, and this is an area of research that deserves more attention. Finally, within the category of household structure variables, single person and single parent households make more stops than nuclear family households, a result consistent with those from other studies (Strathman *et al.*, 1994).

The accessibility of the household residential zone to shopping opportunities did not indicate a statistically significant impact on stop-making propensity when introduced directly in the ORL specification. We next attempted interaction effects of the accessibility measure with indicators for urban, suburban, and rural residential locations. The interaction of the accessibility term with urban and suburban locations were statistically very insignificant, and are therefore not shown in Table 2. The interaction term for rural location indicates the strong impact of accessibility on the propensity to make shopping stops for households residing in rural areas. The MORL model also shows a similar result. However, in addition, the MORL model indicates heterogeneity in the sensitivity of households to shopping accessibility across rural areas. The estimates of the mean and standard deviation on the interaction variable of rural location and accessibility suggest that the effect of accessibility, though varying across rural zones, remains positive for almost all rural zones (the effect of accessibility on household stop-making is negative in less than 4% of zones, as implied by the normal distribution of the coefficient)¹. To summarize, it appears that the impact of accessibility on shopping stop-making propensity may be important only when the existing level of accessibility

¹The normal distribution assumption implies that the effect of accessibility on household stop-making propensity will be negative for some share of rural zones. Alternatively, one can use a log-normal distribution to impose a positive coefficient for all zones. However, the log-likelihood of the resulting model was lower than the one with a normally distributed coefficient.

is quite low. In urban and suburban areas where there is a good existing distribution of opportunities for shopping within close proximity, an increase or decrease in the accessibility measure appears to be inconsequential.

In addition to allowing spatial heterogeneity in the effect of independent variables across zones, the MORL model also accommodates intrinsic differences in stop-making propensity across zones (*i.e.*, intercept unobserved heterogeneity). To allow differences in the extent of this intercept unobserved heterogeneity across zones (*i.e.*, to accommodate spatial heteroscedasticity), we use an exponential form to relate zonal attributes to the standard deviation characterizing the intercept heterogeneity (see section 2). The only two variables found to significantly affect this standard deviation were dummy variable indicators identifying whether a zone was in a suburban or rural area (see bottom of Table 2). The table does not indicate a *t*-statistic for the constant under intercept unobserved heterogeneity because the only logical statistical test for the value of this parameter is with negative infinity, corresponding to the absence of intercept unobserved heterogeneity. However, when zonal attributes are not introduced in the unobserved heterogeneity term, one can directly estimate the fixed variance (without using an exponential functional form) and compare this variance estimate with zero to test for presence of intercept unobserved heterogeneity (maintaining homoscedasticity of the variance parameter). The resulting variance parameter is 0.30 and the *t*-statistic with respect to zero is 3.14. This indicates the presence of statistically significant intercept unobserved heterogeneity. Of course, the MORL model in Table 2 is more general and allows heteroscedasticity in the intercept unobserved heterogeneity. The results indicate more variability in shopping stop-making propensity across suburban and rural zones relative to urban zones.

4.5. Model application

The model estimated in this paper can be used to determine the change in the number of shopping stops due to changes in socio-demographic characteristics over time. This is particularly important because of changing demographic and employment-related trends over the next few decades. For instance, the structure of the household is changing rapidly with an increase in households with no children (Texas State Data Center, 2000). The number of employed individuals in the household are also on the rise (U.S. Bureau of Census, 1996) and this trend is likely to continue. Such demographic and employment changes will have an effect on number of shopping stops, and the model in this paper can be used to assess these impacts and provide reliable information regarding stop-making for input to travel demand and air quality analysis.

The model can also be used to assess the impact of policy actions that involve a change in land-use/transportation level-of-service. These changes can be incorporated by modifying the accessibility variable in response to land-use and transportation level-of-service changes.

In this paper, we demonstrate the application of the model by studying the effect of two changes. The first is an increase in the number of couple family households and the second is an increase in accessibility in rural areas. The increase in couple family households is reflected by randomly selecting 35% of current nuclear family households in the sample and designating them as couple families. As indicated earlier, such a change mirrors the projected decrease in households with no children in the US population. With this change, the number of couple family households in the sample increases from 478 (26.3%) to 685 (37.7%). The change in accessibility in rural areas is “implemented” by increasing accessibility in rural areas by 20%.

The impact of the two changes discussed above is evaluated by modifying exogenous variables to reflect a change, computing revised expected aggregate values for number of trips in

each stop category, and then obtaining a percentage change from the baseline estimates. For the change in number of couple family households, the aggregation is conducted over the entire sample to determine the population change in number of stops. However, for the accessibility change, the aggregation is confined to households located in rural areas, since the results indicate that accessibility does not impact stop-making for households in non-rural areas.

Table 3 provides the results estimated by the (aspatial) ORL model and the (spatial) MORL model. In response to the increase in couple family households, the ORL and MORL models indicate a reduction in the number of households making zero stops and progressively higher percentage increases in the number of households making one, two, three, four, and >4 stops. The reduction in the number of households making zero stops, and the increase in the number of households making more than four stops, is an immediate consequence of the ordered-response structure and the overall positive effect of couple families on shopping stop-making propensity (see Table 2). The positive and progressively increasing number of households making one, two, three, and four stops are not immediate consequences of the overall positive effect of couple families on stop-making propensity, but are empirical results obtained in the current context. Between the ORL and MORL model, the ORL model predicts a smaller percentage decrease in the number of households making zero stops, and smaller percentage increases in the number of households making one, two, three, four and more than four stops. Similar results are obtained from the two models in response to an increase in rural accessibility.

The net percentage change in number of shopping stops can be computed as:

$$\text{Net \% change in shopping stops} = \sum_{k=0}^K \left(\frac{kh_k}{\sum_k kh_k} \right) \theta_k \quad (12)$$

where h_k is the expected number of households who make k stops and θ_k is the percentage aggregate change in each stop category. This overall effect on number of stops is shown in the final column of Table 3. The ORL model underestimates the increase in stops due to an increase in couple family households by as much as 50%. It also underestimates the increase in stops made by rural households due to an increase in rural accessibility by about 63%. Since each shopping stop contributes an additional trip, the incorrect predictions from the ORL model can lead to mis-guided transportation planning strategies and to implementation of transportation control measures.

5. Summary and Conclusions

This paper has proposed a mixed ordered logit model (MORL) for the spatial analysis of household shopping stop-making behavior within the context of a multi-level analysis framework. To the author's knowledge, this is one of the first travel demand modeling studies that recognizes and accommodates spatial considerations in analysis. This is indeed surprising, given that travel-related choices are intrinsically spatial. While GIS platforms are increasingly being used by many metropolitan and state agencies for travel analysis, they are being primarily used for data assembly and presentation of model results. The underlying analytic methods have remained aspatial.

The MORL model formulated in this paper accommodates spatial heterogeneity across traffic analysis zones in the relationship between household stop-making behavior and its determinants. It also captures spatial autocorrelation among households residing within the same zone, and recognizes spatial heteroscedasticity. The MORL model is estimated using a simulated maximum likelihood method with quasi-random Halton draws. The Halton method uses cleverly crafted non-random and uniformly distributed sequences in the domain of integration, and has been shown to be

substantially more efficient than the commonly used pseudo-random method for the mixed logit structure.

The empirical analysis in the paper uses the 1991 Boston area activity survey to examine the determinants of household shopping stop behavior. The empirical results indicate the important effects of employment and household structure on stop-making behavior. The results also suggest that accessibility to shopping opportunities influences the number of shopping stops made, though this effect appears to be confined to rural areas with low existing accessibility. This finding implies that transportation policies such as congestion pricing or peak period pricing will not reduce participation in shopping activities in urban settings.

A comparison of the aspatial and spatial models estimated in the current paper indicates the significant presence of spatial dependency, spatial heterogeneity, and spatial heteroscedasticity in household stop making propensity. Accommodating these spatial issues in the analysis leads to a statistically superior data fit in the empirical context of this study and also provides useful information regarding the differential responsiveness of households across zones.

The aspatial and spatial models are applied to evaluate the effect of a change in the number of couple family households “over time” and the effect of an increase in rural accessibility. The results of this application indicate that the aspatial model underestimates the increase in stops due to an increase in couple family households and an increase in rural accessibility in the empirical context of this paper. In summary, the study underscores the importance of accommodating and testing for the presence of unobserved heterogeneity in the modeling of stop-making decisions.

The current paper contributes to the spatial analysis literature by proposing a multi-level, mixed logit, framework for discrete choice variables. However, one important limitation of the multi-level framework must be acknowledged; the framework requires the *a priori* definition of

traffic analysis zones in the metropolitan region and assumes that the spatial process underlying activity stop generation is influenced in the same way within each zone, but is influenced in an entirely different way as soon as the boundaries of the zones are reached (see Fotheringham and Brunson, 1999). Relaxing this assumption to recognize and accommodate the continuous nature of space is a useful and challenging avenue for future research.

Acknowledgments

This research was funded in part by National Science Foundation grants DMS 9208758 and DMS 9313013 to the National Institute of Statistical Sciences (NISS). The authors are very grateful to Lisa Weyant for her help in typesetting and formatting this document.

References

- Agyemang-Duah, K. and F.L. Hall, (1997). Spatial transferability of an ordered-response model of trip generation. *Transportation Research A* 31, 389-402.
- Anselin, L., Griffith, D.A., 1988. Do spatial effects really matter in regression analysis? *Papers of the Regional Science Association*, 65, 11-34.
- Badoe, D.A., Miller, E.J., 2000. Transportation-land-use interaction: empirical findings in North America, and their implication for modeling. *Transportation Research D*, 235-263.
- Bhat, C.R., 1998. Accommodating variations in responsiveness to level-of-service variables in travel mode choice modeling. *Transportation Research A* 32, 495-507.
- Bhat, C.R., 1999. An analysis of evening commute stop-making behavior using repeated choice observations from a multi-day survey. *Transportation Research B* 33 (7), 495-510.
- Bhat, C. R., 2000a. A multi-level cross-classified model for discrete response variables. *Transportation Research B* 34, 567-582.
- Bhat, C.R., 2000b. Flexible model structures for discrete choice analysis. In: Hensher, D.A., Button, K.J. (Eds), *Handbook of Transport Modelling*. Elsevier, Oxford.
- Bhat, C.R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research B* 35 (7), 677-693.
- Bhat, C.R., Carini, J.P., Misra, R., 1999. Modeling the generation and organization of household activity stops. *Transportation Research Record* 1676, 153-161.
- Bhat, C.R., Govindarajan, A., Pulugurta, V., 1998. Disaggregate attraction-end choice modeling. *Transportation Research Record* 1645, 60-68.
- Bhat, C.R., Koppelman, F.S., 1999. A retrospective and prospective survey of time-use research. *Transportation* 26 (2), 119-139.
- Bhat, C.R., Singh, S.K., 2000. A comprehensive daily activity-travel generation model system for workers. *Transportation Research* 34A (1), 1-22.
- Brownstone, D., Train, K., 1999. Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics* 89, 109-129.
- Case, A., 1992. Neighborhood influence and technological change. *Regional Science and Urban Econometrics* 22, 491-508.

- Dubin, R., 1995. Estimating logit models with spatial dependence. In: Anselin, L. Florax, R.J.G.M. (Eds.), *New Directions in Spatial Econometrics*. Springer, Berlin.
- Fotheringham, A.S., 1983. Some theoretical aspects of destination choice and their relevance to production-constrained gravity models. *Environment and Planning A* 15, 1121-1132.
- Fotheringham, A.S., Charlton, M.E., Brunson, C., 1996. The geography of parameter space: an investigation into spatial non-stationarity. *International Journal of Geographic Information Systems* 10, 605-627.
- Fotheringham, A.S., Charlton, M.E., Brunson, C., 1997. Two techniques for exploring non-stationarity in geographical data. *Geographical Systems* 4, 59-82.
- Fotheringham, A.S., Brunson, C., 1999. Local forms of spatial analysis. *Geographical Analysis* 31(4), 340-358.
- Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., Thomas, S., 1993. A multilevel analysis of school examination results. *Oxford Review of Education* 19, 425-433.
- Goulias, K.G., 2000. Multilevel analysis of daily time use and time allocations to activity types accounting for complex covariance structures using correlated random effects. Presented at the Ninth International Association for Travel Behaviour Research (IATBR) Meeting, Gold Coast, Queensland, Australia, July 2-7.
- Greene, W.H., 1997. *Econometric Analysis*. Prentice Hall, Englewood Cliffs, NJ.
- Hox, J.J., Kreft, I.G., 1994. Multilevel analysis methods. *Sociological Methods and Research* 22, 283-299.
- Hajivassiliou, V.A., Ruud, P.A., 1994. Classical estimation methods for LDV models using simulation. In: Engle, R.F., McFadden, D.L. (Eds.), *Handbook of Econometrics*, vol. 4. North Holland, New York, pp. 2383-2441
- Hensher, D.A., 1999. The valuation of travel time savings for urban car drivers: evaluating alternative model specifications. Technical Paper, Institute of Transport Studies, The University of Sydney, Australia.
- Jones, K., Bullen, N., 1994. Contextual models of urban home prices: a comparison of fixed and random coefficient models developed by expansion. *Econometric Geography* 70, 252-272.

- Jones, K., Duncan, C., 1996. People and places: the multilevel model as a general framework for the quantitative analysis of geographical data. In: Longley, P., Batty, M. (Eds.), *Spatial Analysis: Modelling in a GIS Environment*. GeoInformational International, Cambridge.
- Kitamura, R., Chen, C., Pendyala, R.M., Narayanan, R., 2000. Micro-simulation of daily activity-travel patterns for travel demand forecasting. *Transportation* 27(1), 25-51.
- Langford, I.H., Lewis, T., 1998. Outliers in multilevel data. *Royal Statistical Society* 161(2), 121-160.
- Lee, L.F., 1992. On the efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. *Econometric Theory* 8(4), 518-552.
- LeSage, J.P., 2000. Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis* 32(1), 19-35.
- McKelvey, R.D., Zavoina, W., 1975. A statistical model for the analysis of ordinal-level dependent variables. *Journal of Mathematical Sociology* 4, 103-120.
- McMillen, D.P., 1992. Probit and spatial autocorrelation. *Journal of Regional Science* 32(3), 335-348.
- McMillen, D.P., 1995. Selection bias in spatial econometric models. *Journal of Regional Science* 35(3), 417-438.
- Miller, H. J., 1999. Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T). *Geographical Analysis* 31(4), 373-399.
- Misra, R., Bhat, C.R., 2000. Activity travel patterns of non-workers in the San Francisco Bay area: exploratory analysis. *Transportation Research Record* 1718, 43-51.
- Niederreiter, H., 1995. New developments in uniform pseudo-random number and vector generation. In: Niederreiter, H., Shiue, J.-S. (Eds.), *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, Springer, New York, pp. 94-112,.
- O'Muircheartaigh, C., Campanelli, P., 1999. A multilevel exploration of the role of interviewers in survey non-response. *Royal Statistical Society* 162(3), 437-446.
- Revelt, D., Train, K., 1998. Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics* 80, 647-657.
- Ribaudo, H.J., Bacchi, M., Jurg, B., Thompson, S. G., 1999. A multilevel analysis of longitudinal ordinal data: evaluation of the level of physical performance of women receiving adjuvant therapy for breast cancer. *Royal Statistical Society* 162(1), 349-360.

- Smit, L., 1997. Changing commuter distances in the Netherlands: a macro-micro perspective. places and people: multilevel modeling in geographical research. In: Westert, G.P., Verhoeff, R.N. (Eds.), *Nederlandse Geografische Studies* vol. 227, University of Utrecht, pp. 86-99.
- Strambi, O., van de Bilt, K.A., 1998. Trip generation modeling using CHAID, a criterion-based segmentation modeling tool. *Transportation Research Record* 1645, 24-31.
- Strathman, J.G., Dueker, K.J., Davis, J.S., 1994. Effects of household structure and selected travel characteristics on trip chaining. *Transportation* 21, 23-45.
- Swait, J., Adamowicz, W., 1996. The effect of choice environment and task demands on consumer behavior: discriminating between contribution and confusion. Working paper, Department of Rural Economy, University of Alberta.
- Texas State Data Center & The Center for Demographic and Socioeconomic Research and Education, 2000. Projections of the population of Texas and counties in Texas by age, sex and race/ethnicity for 1990-2030. College Station, Texas A&M University, <http://txsdc.tamu.edu/cgi-bin/prjctn2000.cgi>.
- Train, K., 1999. Halton sequences for mixed logit. Technical paper, Department of Economics, University of California, Berkeley.
- U.S. Bureau of the Census, 1996. Population projections for states by age, sex, race, and Hispanic origin: 1995-2025. Population Projections Branch. <http://www.census.gov/population/www/projections/pp147.html>
- Verheij, R.A., 1997. Physiotherapy utilization: does place matter? In: Westert, G.P., Verhoeff, R.N. (Eds.), *Places and People: Multilevel Modeling in Geographical Research*. *Nederlandse Geografische Studies* 227, University of Utrecht, pp 74-85.
- Washington, S., Wolf, J., 1997. Hierarchical tree-bases versus ordinary least squares linear regression models: theory and example applied to trip generation. Presented at 76th Annual Meeting of the Transportation Research Board. Washington, D.C.
- Wen, C., Koppelman, F.S., 2000. A conceptual and methodological framework for the generation of activity-travel patterns. *Transportation* 27(1), 5-23.
- Yang, M., Goldstein, H., Heath, A., 2000. Multilevel models for repeating binary outcomes: attitudes and voting over the electoral cycle. *Royal Statistical Society* 163(1), 49-62.

LIST OF TABLES

TABLE 1 Descriptive Sample Statistics of Exogenous Variables

TABLE 2 Empirical Results for Household Shopping Stop-Making Propensity

TABLE 3 Impact of Changes on Number of Shopping Stops

TABLE 1 Descriptive Sample Statistics of Exogenous Variables

Variable	Mean	Std. Dev.	Min.	Max.
Household employment related variables				
Number of individuals employed full-time	1.06	0.81	0.00	5.00
Number of individuals employed part-time	0.29	0.54	0.00	3.00
Number of unemployed individuals	0.53	0.72	0.00	4.00
Income x 10 ⁻⁴	4.77	2.78	1.50	11.50
Household structure variables				
Cohabitation	0.09	0.29	0.00	1.00
Couple family	0.26	0.44	0.00	1.00
Single person household	0.31	0.46	0.00	1.00
Single parent family	0.02	0.13	0.00	1.00
Nuclear family and other	0.32	0.53	0.00	1.00
Presence of children < 5 years	0.11	0.31	0.00	1.00
Presence of children between 5 - 11 years	0.07	0.26	0.00	1.00
Presence of children between 12 - 16 years	0.06	0.24	0.00	1.00
Household location variables				
Accessibility measure	1.57	0.14	1.29	2.50
Urban location	0.34	0.47	0.00	1.00
Suburban location	0.21	0.41	0.00	1.00
Rural location	0.45	0.50	0.00	1.00

TABLE 2 Empirical Results for Household Shopping Stop-Making Propensity

Variable ¹	Aspatial ordered response logit (ORL)		Spatial mixed ordered response logit (MORL)	
	Parameter	t-stat.	Parameter	t-stat.
Household employment-related variables				
Number of individuals employed full-time	0.263	2.59	0.244	1.98
Number of individuals employed part-time	0.615	5.38	0.607	3.94
Number of unemployed individuals	0.896	8.67	0.902	6.29
Income x 10 ⁻⁴	0.069	3.51	0.068	2.80
Household structure variables				
Presence of children 12-16 years				
- Mean	0.285	1.50	0.120	0.31
- Std. dev.	-	-	1.150	2.05
Couple and cohabitation families				
- Mean	0.321	2.15	0.201	1.19
- Std. dev.	-	-	0.891	4.07
Single person household	0.451	2.32	0.320	1.23
Single parent family	0.997	2.68	0.892	1.78
Household location variables				
Accessibility x rural location				
- Mean	0.169	2.651	0.204	4.01
- Std. dev.	-	-	0.113	2.21
Intercept unobserved heterogeneity				
Constant	-	-	-1.033	-
Suburban household location	-	-	-0.645	2.30
Rural household location	-	-	-0.485	2.50
Log-likelihood at convergence ²	-2251.25		-2212.62	

¹The threshold values and their standard errors (in parenthesis) for the ORL model are 1.45 (0.27), 2.96 (0.26), 4.18 (0.29), 5.26 (0.31) and 6.16 (0.33). The corresponding values for the MORL model are 1.31 (0.32), 2.92 (0.33), 4.22 (0.34), 5.35 (0.41) and 6.28 (0.43).

²The log-likelihood value at convergence for some other models are as follows: a) with threshold values only (*i.e.*, market shares model): -2331.3, b) ORL model with homoscedastic intercept unobserved heterogeneity: -2235.68, c) ORL model with heteroscedastic intercept unobserved heterogeneity: -2228.23.

TABLE 3 Impact of Changes on Number of Shopping Stops

Change	Model ¹	Percentage aggregate change in shopping stops						Net effect
		0 stops	1 stop	2 stops	3 stops	4 stops	> 4 stops	
Increase in couple family households	ORL	-1.678	0.276	2.276	3.745	4.581	5.069	2.218
	MORL	-1.805	0.153	2.796	5.207	9.976	12.740	3.322
Increase in rural accessibility	ORL	-2.721	0.650	2.889	4.059	4.610	4.918	2.218
	MORL	-3.124	0.489	3.169	5.783	5.854	6.480	3.609

¹ ORL refers to the a spatial ordered-response logit model; MORL refers to the spatial mixed ordered-response logit model.