The Spatial Distribution of Retail Expenditures: Joint Estimation of a Polychotomous Discrete-Continuous Choice System.

Peter O. Barnard
Australian Meat and Livestock Association

and

David A. Hensher
Institute of Transport Studies
Graduate School of Management and Public Policy
University of Sydney

August 12, 1991

ITS-WP-91-7

Acknowledgments

This research was financially supported by the **South Australian Department of Transport, the Australian Road Research Board and the University of Sydney.**

Abstract

A polychotomous discrete-continuous choice system of the spatial distribution of retail expenditures is jointly estimated which allows the imposition of cross-equation restrictions between functions describing discrete and continuous choices as implied by economic theory. The empirical model fuses the shopping destination choices made by individuals with shopping expenditure decisions. The econometric and empirical model offers a rich insight into shopping behaviour and demonstrates the benefits of joint estimation of discrete-continuous choices in contrast to sequential estimation. The approach has wide applicability to many problems involving discrete and continuous choices which are jointly determined.

INTRODUCTION

A growing area of interest in transport economics is the relationship between discrete and continuous choices. Discrete-continuous choice models have been used to study choices in a number of areas such as residential appliance holdings and consumption (Dubin and McFadden (1984) and Brownstone (1980)) and automobile holdings and utilisation (Mannering and Winston (1985), Train (1986), Hensher and Milthorpe (1987), and Hensher et. al. (1992)).

In the present study the economic links between discrete and continuous choices are used to analyse shopping behaviour. The empirical study forges a link between two hitherto disparate approaches to examining shopping behaviour. One approach characterised by discrete choice shopping models analyses the decision of where to shop in isolation of how much to spend; for example the contributions of Domencich and McFadden (1975), Recker and Kostyniuk (1978), Koppelman and Hauser (1978), McCarthy (1979), Gautshci (1981), Weisbrod et.al. (1984), Parcells and Kern (1984) and Eagle (1984). Another set of models has examined shopping expenditure or retail sales patterns largely ignoring how this is related to individual decisions of where to shop; for example, Curhan (1972), Guy (1984) and Morey (1980). To the extent that these two choices are interrelated, these models will be less than complete and the results may be biased. From an information perspective it is beneficial for developers and planners to know both the number of persons using a shopping centre and the expenditure at that centre.

We extend recent work on discrete-continuous choice modelling by jointly estimating a model system where the discrete choice is characterised by a polychotomous choice. In contrast to two-stage methods, joint estimation allows for the imposition of a number of cross-equation restrictions implied by economic theory. As a result of this, the link between economic theory and the empirical model is stronger than in past studies that have used two-stage estimation techniques.

The remainder of the paper is divided into five sections. The following section sets out a general theoretical framework for analysing shopping destination and expenditure choices. This framework is refined in section 3 in order to derive an empirical and jointly estimable destination-expenditure choice system. Section 4 reports results from an empirical application, and section 5 comments on the benefits of our proposed model over the widely used Huff retail expenditure model.

SHOPPING DESTINATION AND EXPENDITURE CHOICE: THEORY

The conventional economic paradigm of self-gratification assumes that an average consumer, q, selects a shopping destination and levels of shopping expenditure, leisure and consumption of other goods as if to maximise utility. The ***average*** consumer may be identified as the main household shopper for an analysis conducted at the household level. A general form of the consumer's utility function from the analyst's perspective is written as:

$$U_q = U_q (G_q, B_{1q}, B_{2q}, ..., B_{N_q q}, Z_q, L_q, \varepsilon_q) \tag{1}$$

where G_q is a vector $(g_{1q}, g_{2q}, ..., g_{N_q q})$ representing consumption of shopping items by consumer q from destinations 1, 2, ..., N_q , respectively, B_{iq} is a vector $(b_{iq1}, b_{iq2}, ..., b_{iqK})$ of K quality variables associated with the consumption of shopping items by consumer q from the ith destination, Z_q is the Hicksian composite commodity encapsulating consumption of other goods, L_q is leisure time and ε_q is a vector $(\varepsilon_{1q}, \varepsilon_{2q}, ..., \varepsilon_{N_q q})$ of analyst unobserved influences on utility.

Maximisation of utility is subject to income and time constraints:

$$Y_q = \sum p_{iq} \qquad \sum_i^{N_q} \quad c_{iq} \tag{2}$$

$$L_q = T - \sum_i^{N_q} \xi_{iq} t_{iq} \tag{3}$$

where p_{iq} is an index of shopping prices facing individual q at the ith destination, $\xi_{iq} = \xi_{iq} (g_{iq})$ is an indicator function with $\xi_{iq} = 1$ if $g_{iq} > 0$ and $\xi_{iq} = 0$ if $g_{iq} = 0$, c_{iq} is the cost of travel to the ith destination, Y_q is income and T is total time available. In equation [2] income, shopping prices and travel costs have been normalized by the price, P_z, of the Hicksian composite commodity. Alternative i is strictly a destination/mode combination, since travel times and costs vary by alternative modes as well as destinations. It is conceptually easier however to think of i solely in terms of destination choice.

An element of discreteness can be introduced into the model by assuming that in any time period the consumer selects one destination for shopping purchases. A possible behavioural source for this restriction is that the consumer views alternative shopping destinations as perfect substitutes, but one destination must be chosen since shopping represents an essential activity. This effectively concentrates attention on the destination choice of shopping travel behaviour. It implies that Z, L, and one of the g_j's is positive with all $g_i\ (i \neq j)$ equal to zero. The discrete element of the solution relates to which of the g_is are to take zero values. A continuous dimension is also evident because the non-zero g_i, Z and L can be consumed in any quantities.

In obtaining optimal values of the g_is, Z and L the consumer can be thought of as applying a two stage maximisation process. Assuming that shopping destination 1 is chosen, and if $g_{iq} = 0$ then $\frac{\partial U_q}{\partial b_{i1q}} = \frac{\partial U_q}{\partial b_{i2q}} = ... = \frac{\partial U_q}{\partial b_{iKq}} = 0$ (Hanemann 1984), the maximisation problem can be redefined as:

$$\max U_{1q} = U_{1q}(\ g_{1q},\ B_{1q},\ Z_q,\ L_q,\ \varepsilon_q\ ) \tag{4}$$

subject to:

$$Y_q = p_{1q}\, g_{1q} + Z_q + c_{1q} \tag{5a}$$

$$L_q = T - t_{1q} \tag{5b}$$

The solution to (4), (5a), (5b) is a set of demand equations conditional upon the choice of shopping destination 1:

$$g^*_{1q} = g^*_{1q}(p_{1q}, B_{1q}, T - t_{1q}, Y_q - c_{1q}, \varepsilon_q) \tag{6a}$$

$$Z^*_{1q} = Z^*_{1q}(p_{1q}, B_{1q}, T - t_{1q}, Y_q - c_{1q}, \varepsilon_q) \tag{6a}$$

$$L^*_{1q} = g^*_{1q}(p_{1q}, B_{1q}, T - t_{1q}, Y_q - c_{1q}, \varepsilon_q) \tag{6a}$$

This process is repeated for $g_{2q} > 0,\ g_{1q} = g_{3q} = ... = g_{N_q q} = 0$ and so on. The second stage involves the consumer computing the point of global utility maximisation as:

$$\tilde{U}\,(\tilde{g}^*, \tilde{Z}^*, \tilde{L}^*, \tilde{B}, \varepsilon_q) = \max \left\{ \begin{matrix} U_{1q}\left(g^*_{1q}, Z^*_{1q}, L^*_{1q}, B_{1q}, \varepsilon_q\right), U_{2q}\left(g^*_{2q}, Z^*_{2q}, L^*_{2q}, B_{2q}, \varepsilon_q\right), \\ ..., \ U_{N_q q}\left(g^*_{N_q q}, Z^*_{N_q q}, L^*_{N_q q}, B^*_{N_q q}, \varepsilon_q\right) \end{matrix} \right\} \tag{7}$$

Equation (7) can be alternatively expressed as:

$$\tilde{V}(\tilde{p}, \tilde{B}, T - \tilde{t}, Y_q - \tilde{c}, \varepsilon_q) = \max \{ V_{1q}(p_{1q}, B_{1q}, T - t_{1q}, Y_q - c_{1q}, \varepsilon_q),$$

$$V_{2q}(p_{2q}, B_{2q}, T - t_{2q}, Y_q - c_{2q}, \varepsilon_q), \ldots,$$

$$V_{N_q q}(p_{N_q q}, B_{N_q q}, T - t_{N_q q}, Y_q - c_{N_q q}, \varepsilon_q) \} \tag{8}$$

where V is the indirect utility function. The conditional indirect utility function (CIUF) associated with shopping destination i, V_{iq}, is defined by:

$$V_{iq} = U_{iq}(g^*_{iq}, Z^*_{iq}, L^*_{iq}, B_{iq}, \varepsilon_q) = V_{iq}(p_{iq}, B_{iq}, T - t_{iq}, Y_q - c_{iq}, \varepsilon_q) \tag{9}$$

Shopping destination j will be chosen if:

$$V_{jq}(p_{jq}, B_{jq}, T - t_{jq}, Y_q - c_{jq}, \varepsilon_q) > V_{iq}(p_{iq}, B_{iq}, T - t_{iq}, Y_q - c_{iq}, \varepsilon_q) \text{ for all } i \neq j \tag{10}$$

The V_{jq} are the functions encountered in conventional derivations of discrete choice models (e.g. McFadden 1981, Small 1982, Domencich and McFadden 1975, Hensher and Johnson 1981, Greene 1990).

For certain functional forms, in comparing shopping destinations i and j, T and Y_q can be deleted. These functions contain variables describing prices at destination i, other attractiveness variables associated with destination i, and travel times and costs to destination i - all the variables normally included in a behaviourally based shopping destination choice model (Recker and Kostyniuk 1978, Koppelman and Hauser 1977, McCarthy 1979, Gautschi 1981, Weisbrod et.al. 1984, Parcells and Kern 1984, Eagle 1984).

The convenience of working with the indirect utility function derives from the knowledge that demand equations which are consistent with utility maximising behaviour can be obtained by applying Roy's identity (Roy 1942) to V_{iq}, rather than explicitly solving the maximisation problem in equations (4) and (5). In particular, the conditional demand equation corresponding to the CIUF shown in equation (9) can be derived as:

$$g^*_{iq} = - \frac{\partial V_{iq}/\partial p_{iq}}{\partial V_{iq}/\partial Y_q} = g^*_{iq}(p_{iq}, B_{iq}, T - t_{iq}, Y_q - c_{iq}, \varepsilon_q) \tag{11}$$

Equation (11) can be expressed in expenditure form as:

$$p_{iq}g^*_{iq} = E_{iq} = p_{iq}g^*_{iq}\,(p_{iq}\,,\, B_{iq}\,,\, T - t_{iq}\,,\, Y_q - c_{iq}\,,\, \varepsilon_q\,) \tag{12}$$

Equation (12) and the corresponding indirect utility function (equation (9)) establishes the link between shopping centre choice and shopping expenditure.

ISSUES IN THE EMPIRICAL ESTIMATION OF A JOINT SHOPPING DESTINATION AND EXPENDITURE CHOICE MODEL

In deriving an empirically estimable model we distinguish between the observable and unobservable components of the CIUFs and conditional expenditure functions by specifying:

$$V_{iq} = \overline{V}_{iq}\,(\,p_{iq}\,,\, B_{iq}\,,\, T - t_{iq}\,,\, Y - c_{iq},\,) + \varepsilon_{iq} \tag{13}$$

and

$$E_{iq} = p_{iq}\overline{g}^*_{iq}\,(\,p_{iq}\,,\, B_{iq}\,,\, T - t_{iq}\,,\, Y - c_{iq},\,) + u_{iq} \tag{14}$$

where $\overline{V}_{iq}$ is the observable or representative component of the CIUFs, $\overline{g}^*_{iq}$ is the representative component of the conditional demand functions and ε_{iq} and u_{iq} are error terms in the CIUFs and expenditure functions, respectively.

To simplify computational aspects of the model system, we have chosen to specify a form for the $\overline{V}_{iq}$ that will yield, after application of Roy's identity, a linear-in-the-parameters shopping expenditure model. A family of CIUFs that meet this requirement is defined by:

$$\overline{V}_{iq} = f_1 \left\{ f_2 \left[\log p_{iq}, B_{iq}, Y - c_{iq}, T - t_{iq}\right] p_{iq}^{-\alpha_e} \right\} \tag{15}$$

where f_2 is linear in its arguments. A specific form of equation (15) is:

$$\overline{V}_{iq} = \left[\alpha_1 - \alpha_2 \log(p_{iq}) + \sum_{k=3}^{K+2} \alpha_k b_{iqk} + \alpha_{K+3} (Y_q - c_{iq}) + \alpha_{K+4} (T - t_{iq}) \right] p_{iq}^{-\alpha_{K+3}} \tag{16}$$

which is the form for the 'representative' component of the CIUFs utilized in the current study. Equation (16) is a variant of the form of CIUF used by Dubin and McFadden (1984) in a binary choice context. Applying Roy's identity to equation (16) provides the expected shopping expenditure level for consumer q:

$$\overline{E}_{iq} = \alpha_1 - \alpha_2 \log(p_{iq}) + \sum_{k=3}^{K+2} \alpha_k b_{iqk} + \alpha_{K+3} (Y_q - c_{iq}) + \alpha_{K+4} (T - t_{iq}) + \frac{\alpha_2}{\alpha_{K+3}} \tag{17}$$

For equation (13) to represent a valid CIUF, with $\overline{V}_{iq}$ defined by equation (16), it must conform to a number of conditions (Diewert 1974):

(i) V(.) is continuous for all prices and income > 0,

(ii) V(.) is homogeneous of degree zero in prices and income,

(iii) V(.) is non-increasing in prices and non-decreasing in income,

and

(iv) V(.) is quasi-convex in prices.

Condition (ii) is automatically met by the formulation of the model. The other conditions are tested upon estimation of the model.

The specification adopted in equations (13) and (14) implies treatment of the unobservable components in the CIUFs and conditional expenditure functions outside the strict theoretical framework. A statistical procedure is now outlined for treating the unobservable components that is in harmony with the theory developed above. In presenting the statistical procedure we economise on notation by defining a row vector Z_{iq} containing $\log(p_{iq})$, B_{iq}, $T - t_{iq}$ and $Y_q - c_{iq}$. Equations (13), (16), and (14), (17) can then be rewritten as:

$$V_{iq} = \overline{V}_{iq}(Z_{iq}, \alpha) + \varepsilon_{iq} \tag{18}$$

and,

$$E_{iq} = Z_{iq}\,\beta + u_{iq} \tag{19}$$

where α and β are the unknown parameter vectors in the discrete and continuous choice models respectively, with $\beta_e = f_e\,(\alpha)$.

There are two important features of equation system (18) and (19). Firstly, shopping expenditure for each individual is only observed at the chosen destination. From the theoretical model, in choosing a shopping destination the individual calculates optimal expenditure levels at each destination; but the optimal expenditure levels at non-chosen centres are hidden from the analyst. The result is systematic missing data on the E_{iq} . Secondly, it is likely that the error terms ε_{iq} and u_{iq} will be correlated because the disturbances in both the discrete destination choice model and continuous shopping expenditure model arise from the same source, namely, uncertainty concerning the CIUFs.

In accordance with most past studies of shopping destination choice, it is assumed that the ε_{iq} are independently and identically distributed extreme value type 1, leading to a multinomial logit (MNL) destination choice model, and the u_{iq} are normally distributed. Recognising the conditionality of observed data points in the expenditure model, equation (19) may be respecified in estimation form as:

$$E_{jq} = Z_{jq}\beta + E\left(u_{jq} \mid \overline{V}_{jq} + \varepsilon_{jq} > \overline{V}_{iq} + \varepsilon_{iq} \text{ for all } i \neq j\right) \tag{20}$$

where E ( ) denotes 'the expected value of '. The last term on the RHS of equation (20) will, in general, be non-zero resulting in biased estimates of β when using OLS.

Using a technique developed by Lee (1983), building upon the work of Heckman (1976), the term $E\left(u_{jq} \mid V_j > V_i \text{ for all } i \neq j\right)$ can be evaluated by the following method. Let,

$$\eta_{jq} = \max\left[V_{iq}\right] - \varepsilon_{jq}\left(i = 1, 2, ..., \; N_q, i \neq j \right) \tag{21}$$

Shopping destination j will be chosen by consumer q if $V_{jq} > V_{iq}$ for all $i \neq j$ or,

$$\tau_q = j \text{ iff } \eta_{jq} < \overline{V}(Z_{jq}, \alpha) \tag{22}$$

With the ε_{iq} iid extreme value type 1, the distribution of η_{jq} is :

$$D(\eta_{jq}) = \exp(\eta_{jq}/\mu) \Big/ \left[\exp(\eta_{jq}/\mu) + \sum_{\substack{i=1 \\ i \neq j}} \exp\left[\overline{V}(Z_{iq}, \alpha)/\mu\right]\right] \tag{23}$$

where μ is the scale parameter of the logit distribution $\left(\mu = (\sqrt{3}\ \sigma_{\varepsilon_i\varepsilon_i})/\pi \text{ with } (\sigma_{\varepsilon_i\varepsilon_i})^2 \text{ the variance of } \varepsilon_i\right)$. In turn, η_{jq} can be transformed into a standard normal variate, η^*_{jq} by applying:

$$\eta^*_{jq} = J(\eta_{jq}) = \phi^{-1}\left[D(\eta_{jq})\right] \tag{24}$$

where ϕ^{-1} is the inverse of the standard normal distribution. Computationally accurate methods are available for approximating the inverse of the standard normal distribution (NAG 1984). With this transformation j will be chosen iff $\eta^*_{jq} < J\left[\overline{V}(Z_{jq}, \alpha)\right]$.

Given that the u_{iq} are also normally distributed, the bivariate distribution between η^*_j and u_j can be specified as $N\left(0, 0, 1, \sigma_{u_ju_j}, \rho_{u_j\eta^*_j}\right)$. The equation system (18) and (19) should only be estimated independently when the correlation coefficient, $\rho_{u_j\eta^*_j}$, is equal to zero. In the more general case, the conditional expectation $E(u_{jq} \mid I_q = j)$ needs to be included as a regressor in equation (19). Through integration this can be shown to equal:

$$\sigma_{u_j\eta^*_j}\,\Phi\left[J\left[\overline{V}(Z_{jq}, \alpha)\right]\right] / D\left[\overline{V}(Z_{jq}, \alpha)\right] \tag{24a}$$

where $\sigma_{u_j\eta^*_j}$ is the covariance between u_j and η^*_j and Φ is the density function of the standard normal, so the shopping expenditure model becomes:

$$E_{jq} = Z_{jq}\beta - \sigma_{u_j\eta^*_j}\,\Phi\left[J\left[\overline{V}(Z_{jq}, \alpha)\right]\right] / D\left[\overline{V}(Z_{jq}, \alpha)\right] + \vartheta_{jq} \tag{25}$$

where ϑ_{jq} is a new error term with $E\left(\vartheta_{jq} \mid I_q = j\right) = 0$.

A two-stage procedure for estimating the model system implied by equations (22) and (25) is:

1. Estimate an MNL model of the choice of shopping destination, obtaining values for $\hat{\alpha}$ and $\hat{\mu}$

2. For the chosen store calculate :

$$J\left[\overline{V}\left(Z_{jq} , \hat{\alpha}\right)\right] = \phi^{-1}\left[D\left[\overline{V}\left(Z_{jq} , \alpha\right)\right]\right] = \phi^{-1}\left(\hat{P}_{jq}\right) \text{ and } D\left[\overline{V}\left(Z_{jq} , \hat{\phi}\right)\right] = \hat{P}_{jq} ,$$

3. Estimate equation (25) using OLS with $\hat{\sigma}_{u_j \eta_j^*}$ being the parameter estimate for selectivity correction, and

4. Correct the variance/covariance matrix associated with the OLS estimation of equation (25). This correction is necessary because the ϑ_{jq} are heteroskedastic. Correction formulae for Lee's selectivity correction method when $\overline{V}$ is non-linear in the parameters are derived in Barnard (1987).

Alternatively the system can be estimated by full information maximum likelihood (FIML). The log likelihood function for a sample of Q individuals is:

$$\log L = \sum_{q}^{Q} \sum_{i}^{N_q} \Big[k_{iq}$$

$$+ k_{iq} \log \phi \ \left[\ \left\{ \ \phi^{-1} \left[D \left(\overline{V}_{iq}\right) \right] - \rho_{u_i \eta_i^*} \left(u_{iq} / \sigma_{u_i u_i}\right) \right\} \ / \ \left(1 - \rho^2_{u_i \eta_i^*}\right)^{0.5} \ \right] \ \Big] \tag{26}$$

where $k_{iq} = 1$ iff $I_q = i$. The log likelihood function is a member of the set of log likelihood functions considered by Lee (1983). The first partial derivatives of equation (26) with respect to the structural parameters, α, are:

$$\frac{\partial \log L}{\partial \alpha_e} = \sum_{q}^{Q} \sum_{i}^{N_q} \Big[- k_{iq} \frac{u_{iq}}{(\sigma_{u_i u_i})^2} \frac{\check{Z} u_{iq}}{\check{Z} \alpha_e} + k_{iq} \frac{\Phi(\kappa_{iq})}{\phi(\kappa_{iq})}$$

$$\times \Big[\big[\Phi [D (\overline{\nabla}_{iq})] \big]^{-1} \big[D (\overline{\nabla}_{iq}) \Big\{ \frac{\partial(\overline{\nabla}_{iq})}{\partial \alpha_e} - \sum_{j} D (\overline{\nabla}_{jq}) \frac{\partial(\overline{\nabla}_{jq})}{\partial \alpha_e} \Big\} \big]$$

$$- \frac{\rho_{u_i \eta_i^*}}{\sigma_{u_i u_i}} \frac{\partial u_{iq}}{\partial \alpha e} \Big] \Big/ [1 - (\rho_{u_i \eta_i^*})^2]^{0.5} \Big] \qquad (27)$$

where $\kappa_{iq} \equiv \Big[\phi^{-1} [D (\overline{\nabla}_{iq})] - \rho_{u_i \eta_i^*} (u_{iq} / \sigma_{u_i u_i}) \Big] / [1 - (\rho_{u_i \eta_i^*})^2]^{0.5}$

Equation (26) can be maximized using a number of algorithms, including the Davidon-Fletcher-Powell algorithm.

EMPIRICAL RESULTS

The model of equation (26) was used to study the distribution of grocery shopping expenditure across stores. The data obtained from a survey of shoppers in Adelaide (Australian Road Research Board, 1981), consisted of two parts. First participating households were required to fill out diaries documenting one week of activities. At the end of that period, diaries were collected and the main household shopper interviewed regarding the household's food shopping arrangements.

In the shopping interview, information was sought on outlets patronized by the household for grocery shopping, and possible alternatives to those outlets. This information defined the household's grocery shopping choice set. For each reported shopping outlet, respondents were asked to rate the outlet in terms of price, selection and store convenience. Selection and store convenience ratings were measured on a 5 point scale with a value range from 'far above average' (5) to 'far below average' (1). The price rating is based on a basket of commonly purchased grocery shopping goods (Choice 1981) and expressed in monetary units.

In addition, information on possible methods of travel to each outlet by the respondent was obtained, and for each mode specified, travel time and cost data collected.

Importantly the same store codes were used in the activity diaries as in the shopping questionnaire. Blank diary pages were divided into two parts. The lower half was designed to facilitate personal documentation of the nature, time, location and level of expenditure associated with each activity episode. The upper half was designed to allow the respondent to provide further information on each trip undertaken (i.e. travel activity).

Data for estimating the model system was obtained from merging the shopping questionnaire information with shopping episodes recorded in the activity diaries. Diaries for main household shoppers who filled in the shopping questionnaire were interrogated for records of activity episodes involving grocery shopping with associated travel to/from the household's residence. Records were rejected if no expenditure information was provided, or if the store visited was not one of the set of stores provided by the shopper in the shopping questionnaire. Further records were excluded when no income information was provided. The estimation data set consisted of observations on 236 store choices. In each case the choice set for an individual comprised the list of mode/store alternatives provided in the shopping questionnaire with the chosen alternative being the mode/store combination observed in the activity diary. Definitions of variables used in this study, along with summary statistical information, are shown in Table 1.

Results from maximising the log likelihood function of equation (26), with $B'_{iq} \equiv (SEL_{iq}, CONV_{iq})$ using the Davidon-Fletcher-Powell algorithm are included in Table 2. A number of initial starting values for the unknown parameters were used to ensure that the global maximum of the log likelihood function had been attained. This procedure is superior to generating initial parameter estimates from two-stage estimation when $\overline{V}_{iq}$ is non-linear in the parameters since the MNL log-likelihood function may be characterized by multiple local maxima. This study and Kristnamurthi and Raj (1988) represent the only known applications of FIML to the joint estimation of a polychotomous discrete/continuous choice model system. McFadden et.al. 1986 is an example of joint estimation of a dichotomous discrete/continuous model system.

TABLE 1

VARIABLE DEFINITIONS AND SUMMARY STATISTICS

Mnemonic	Definition	Mean Value	Standard Deviation
$PRICE_{dq}$	An individual rating of prices at store d based on a basket of commonly purchased grocery shopping goods ($).	25.91	4.88
SEL_{dq}	An individual rating on a 5 point scale of the selection of grocery items available at store d.	2.76	0.81
$CONV_{dq}$	An individual rating on a 5 point scale of the convenience of using store d.	2.22	1.01
$TCOST_{mdq}$	The cost of individual q travelling to and from store d by mode m: - if mode is bus, TCOST = reported bus fare, - if mode is car, TCOST = network highway distance x 0.23, - if mode is walk or bicycle, TCOST = 0 ($).	1.06	0.66
$TTIME_{mdq}$	The reported time for individual q to travel to store d by mode m (minutes).	7.47	4.25
$INCOME_{q}$	Weekly household income ($).	288.96	114.61
$EXPEND_{mdq}$	Grocery shopping expenditure by household q at store d when using mode m.	15.70	17.35

TABLE 2
JOINT STORE CHOICE / SHOPPING EXPENDITURE MODEL PARAMETER ESTIMATES AND STATISTICS

Variable/Parameter	Description	Parameter Estimate	T-Value
CONSTANT	α_1	-13.0779	-10.76
log (PRICE)	α_2	0.7012	3.02
SEL	α_3	0.3758	2.94
CONV	α_4	0.2070	1.79
(INCOME - TCOST)	α_5	0.0469	17.51
(60 - TTIME)	α_6	0.1212	4.19
μ		0.2421	3.63
ρ		0.3357	8.26
σ		16.8174	11.47
Sample Size	236		
log L at convergence	-986.6		
log L (0)*	-1825.6		

* Note: log L(0) is defined as the value of the log-likelihood function with $\alpha_1 = \alpha_2 = ... = \alpha_6 = \rho = 0$ and $\mu = 1$.

In interpreting the parameter estimates of Table 2, from equations (16) and (17), the model specification includes the price related parameter estimates as negatively signed. A positive estimate for α_2 therefore, indicates that as log (p_{iq}) increases, V_{iq} decreases. Similarly, a positive estimate for α_5 indicates as p_{iq} increases, V_{iq} decreases. A number of factors serve to engender confidence in the model. These can be grouped under two headings; compatibility of the results with economic theory and reasonableness of the parameter estimates. With respect to the former, the estimated CIUFs conformed to all necessary properties of indirect utility functions. The non-increasing price condition implies,

$$\frac{\partial V_{iq}}{\partial p_{iq}} = - p^{-(\alpha_5 + 1)} \Big\{ \alpha_5 \Big[\alpha_1 - \alpha_2 \log p_{iq} + \alpha_3 \, SEL_{iq} + \alpha_4 \, CONV_{iq} + \alpha_5 (Y_q - c_{iq}) + \alpha_6 (T - t_{iq}) \Big] + \alpha_2 \Big\} \leq 0 \tag{28}$$

In the sample this condition was met for CIUFs associated with all alternatives in all choice sets. The non-decreasing income condition implies,

$$\frac{\partial V_{iq}}{\partial Y_q} = \alpha_5 \, p^{-\alpha_5} \geq 0 \tag{29}$$

and since $\hat{\alpha}_5 \geq 0$ this condition was also met for all estimated CIUFs. Finally, a test of the quasi-concavity condition is that the diagonal elements of the Slutsky matrix be non-positive (e.g. Hausman 1981). That is,

$$s_{ii} = \frac{\partial Y^*_{iq}}{\partial p_{iq} \partial p_{iq}} = \Big\{ (\alpha_5 - 1) \Big[\alpha_1 - \alpha_2 \log p_i + \alpha_3 \, SEL_{iq} + \alpha_4 \, CONV_{iq} + \alpha_5 (Y_q - c_{iq}) + \alpha_6 (T - t_{iq}) \Big] - \alpha_2 / \alpha_5 \Big\} / p^2_{iq} \leq 0 \tag{30}$$

where Y^*_{iq} is the conditional cost function. In contrast to some other studies which experienced difficulties in meeting this condition (e.g. Wales and Woodland 1977, Brownstone

1980), all estimated CIUFs satisfied equation (30). Evaluated at mean levels for the independent variables

$$\partial V_{iq} / \partial p_{iq} = -\,0.0326, \quad \partial V_{iq} / \partial Y_{q} = 0.0403 \text{ and } s_{ii} = -\,0.0308.$$

The influence of each variable on store choice and grocery shopping expenditure can be gauged from an examination of the relevant elasticities, shown in Table 3. The elasticities with respect to store choice are measured by:

$$\omega_{(dc)\,e} = \sum_{q} \sum_{d} \left[\frac{\partial P_{dq}}{\partial z_{ldq}} z_{edq} \right] / \sum_{q} \sum_{d} P_{dq} \tag{31}$$

TABLE 3
ELASTICITY ESTIMATES FOR STORE CHOICE AND LEVEL OF SHOPPING EXPENDITURE

Variable	Store Choice Elasticities ($\omega_{(dc)}$)	Store Specific Expenditure Elasticities ($\omega_{(x1)}$)	Unconditional Grocery Shopping Expenditure Elasticities ($\omega_{(x2)}$)
PRICE	-0.274	-0.312	-0.038
SEL	0.597	0.630	0.039
CONV	0.277	0.340	0.037
TCOST	-0.042		
TTIME	-0.358		
INCOME			0.717
TIME			0.892

where the subscript d refers to a particular store. Two types of expenditure (= x) elasticities are shown. One is indicative of the impact of a change in a variable on expenditure at a particular store and is calculated as:

$$\omega_{(x1)\,l} = \sum_{q} \sum_{d} \left[\frac{\partial (\bar{E}_{dq} P_{dq})}{\partial z_{edq}} \right] / \sum_{q} \sum_{d} \bar{E}_{dq} P_{dq} \tag{32}$$

where $\widetilde{E}_{dq}$ includes the factor for selectivity correction. The other is indicative of a change in a variable on shopping expenditure in general,

$$\omega_{(x_2)\,l} = \sum_q \left[\beta_e \; z_{edq} \right] / \sum_q E_{dq} \tag{33}$$

Examining the store choice elasticities, as anticipated, an increase in travel time or travel cost to a store results in a decreased probability of that store being selected, as does an increase in store prices. Conversely, an increase in the perceived quality of a destination, as encapsulated in the variables SEL and CONV, is predicted to increase patronage of that destination.

A comparison of the store specific expenditure elasticities with the store choice elasticities, suggests that a change in the value of a variable is predicted to affect store expenditure in the same direction as store patronage, but with greater force. This result has a basis in theory. An increase in prices at a particular store, for example, will not only cause the utility associated with that store to decrease, and thus the probability of choosing the store, but will also cause those individuals who continue to use the store to spend less there.

The unconditional grocery shopping expenditure elasticities possess the same signs as the store specific expenditure elasticities, but are of significantly lesser magnitude. Expenditure on groceries is predicted to be virtually unaffected by a change in grocery prices, in the perceived selection of grocery items available, or in the perceived convenience of using grocery stores. An across-the-board 10% increase in household incomes is predicted to result in a 7.2% increase in grocery shopping expenditure. The time elasticity may be interpreted as the change in expenditure expected if more time for shopping were available.

All the elasticity estimates are of the expected magnitude. The store choice accessibility related elasticity estimates are within the range suggested by other studies which have examined shopping destination choice behaviour (e.g. Domencich and McFadden 1975, Richards and Ben-Akiva 1975). The income inelasticity of grocery expenditures as found in the current study conforms with similar income inelastic estimates for food expenditures obtained in classical studies of consumer demand using substantially different data and statistical methods (Houthakker 1957, Barten 1964, Theil et al. 1981, Podder 1971). Store specific expenditures are predicted to be more sensitive to the perceived range of merchandise available than to prices, an intuitively appealing result. Generally, however, there is a dearth of published estimates of grocery shopping expenditure store attribute elasticities with which to compare the values in column 2 of Table 3.

THE HUFF RETAIL MODEL

A by-product of the analysis is a demonstration of a theoretical inconsistency in a widely used model in retailing, the Huff model. A generalised expression for the Huff model is:

$$C_{qj} = P_{qj} E_q = \sum_i \frac{F_{qj}(D_{qj},\ A_{qj})}{F_{qi}(D_{qi},\ A_{qi})} W(S_q) \qquad (34)$$

where, C_{qj} is the expected expenditure by consumer q at retail trade centre j, P_{qj} is the probability that consumer q will choose to shop at retail trade centre j, E_q is the expenditure on retail goods by consumer q, $A_{q1}, A_{q2}, ..., A_{qN}$ are vectors of variables measuring the attractiveness of shopping centres 1, 2, ..., N to consumer q, $D_{q1}, D_{q2}, ..., D_{qN}$ are vectors of variables measuring the separation of consumer q from shopping centres 1,2, ..., N, S_q is a vector of socioeconomic variables pertaining to consumer q, $F_{q1}, F_{q2}, ..., D_{qj}, ..., F_{qN}$ are functions relating the separation and centre attribute variables to the choice probabilities, and W is a function relating socio-economic characteristics to consumer expenditure levels. Many users of this model have advocated a utility interpretation for the F_{qi} functions. Indeed this suggestion is evident in Huff's original work (Huff 1963).

The treatment of expenditure in Huff-type shopping models is in general inconsistent with a utility based interpretation for the F_{qi} functions. Empirical Models consistent with utility maximisation theory have been derived in previous sections.

It is evident from equation (12) that for a shopping expenditure model to be consistent with utility maximisation expenditure levels must in general be allowed to vary by attributes pertaining to shopping destinations, as well as across individuals. The Huff model, however, only includes socio-economic variables in the expenditure function. Furthermore, when modelling both shopping destination and expenditure choices, the functional form for the expenditure model should bear a direct relationship with the functional form specified for the destination patronage model. As originally specified, and in subsequent applications, shopping expenditure in the Huff model has been treated independently from the F_{qi} functions. Since the Huff model meets neither criteria, it is generally inconsistent with an assumption of utility maximisation.

As a result of this separation, although referring to retail trade area sales potential, most applications of the Huff model have only estimated the patronage probability component of equation (34) (e.g. Stanley and Sewall 1976, Nevin and Houston 1980). This is because for an homogeneous group of consumers, and a particular shopping category, $W(E_q)$ can be regarded as a constant. Due to this practice it has become common to associate the Huff Model with only the patronage probability component of equation (34).

CONCLUSION

In this paper a model which fuses the shopping destination choices made by individuals with shopping expenditure decisions was developed. Economic theory was used to demonstrate a close relationship between these two facets of shopping behaviour. The form of this relationship was then used to develop an empirical model of shopping behaviour.

The use of FIML to jointly estimate the models associated with the store choice and shopping expenditure decisions meant that a set of cross equation parameter restrictions implied by theory could be imposed on estimation. By basing the empirical model of shopping behaviour firmly on economic theory, a number of tests could be applied that are unavailable when ad hoc approaches are used.

Although work reported in this paper has specifically involved the analysis of shopping behaviour, the methods used are applicable to other choice processes, such as those examined by Dubin and McFadden (1984) and Brownstone (1980) or Train (1986), Mannering and Winston (1985) and Hensher et. al. (1992), where a discrete and a continuous component is evident.

REFERENCES

AUSTRALIAN ROAD RESEARCH BOARD (1981). The Adelaide travel demand and time allocation study: questionnaire forms, interview and coding manuals. Australian Road Research Centre, Vermont, Victoria, 121pp.

BARNARD, P.O. (1987). Modelling Shopping Destination Choices: A Theoretical and Empirical Investigation. Australian Road Research Board Special Report No 36, Melbourne.

BARTEN, A.P. (1964). Consumer demand functions under conditions of almost additive preferences. *Econometrica*, 32, 1-38.

BROWNSTONE, D. (1980). An Econometric Model of Consumer Durable Choice and Utilization Rate. Ph.D. Dissertation, University of California, Berkley.

CHOICE (1981). Annual supermarket survey. *Journal of the Australian Consumer's Association*, 11, 171-77.

CURHAN, R.C. (1972). The relationship between shelf space and unit sales in supermarkets. *Journal of Marketing Research*, 9, 406-12.

DIEWERT, W.E. (1974). Applications of duality theory. Chapter 3 in M.D.INTRILIGATOR and D.A. KENDRICK (eds.), *Frontiers of Quantitative Economics*, Vol 11, North Holland, Amsterdam.

DOMENCICH, T.A. and McFADDEN, D. (1975). *Urban Travel Demand*, North Holland, Amsterdam.

DUBIN, J. and McFADDEN, D. (1984). An econometric analysis of residential electric appliance holdings and consumption. Econometrica, 52, 345-62.

EAGLE, T.C. (1984). Parameter stability in disaggregate retail choice models - experimental evidence. *Journal of Retailing*, 60, 101-23.

GAUTSCHI, D.A. (1981). Specification of patronage models for retail centre choice. *Journal of Marketing Research*, 18, 162-74.

GREENE, W.H. (1990) *Econometric Analysis*, MacMillan, New York.

GUY, C.M. (1984). The estimation of retail turnover for planning purposes. *The Planner*, 70, 12-14.

HANEMANN, W.M. (1984). Discrete/continuous models of consumer demand. *Econometrica*, 52, 541-561.

HAUSMAN, J.A. (1981). Exact consumer's surplus and deadweight loss. *American Economic Review*, 71, 662-79.

HECKMAN, J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5, 475-92.

HENSHER, D.A. and JOHNSON, L.W. (1981) Applied Discrete-Choice Modelling, Croom Helm, London and John Wiley, New York.

HENSHER, D.A., SMITH, N.C., MILTHORPE, F.W. and BARNARD, P.O. (1992). *Dimensions of Automobile Demand*: *A Longitudinal Study of Household Automobile Ownership and Use*. North Holland, Amsterdam.

HENSHER, D.A. and MILTHORPE, F. (1987). Selectivity correction in discrete/continuous choice analysis: with empirical evidence for vehicle choice and use. *Regional Science and Urban Economics*, 17, 123-150.

HOUTHAKKER, H. (1957). An international comparison of household expenditure patterns commemorating the centenary of Engel's law. *Econometrica*, 25, 532-551.

HUFF, D.L. (1963). A probabilistic analysis of shopping center trade areas. *Land Economics*, XXXIV, 81-90.

KOPPELMAN, F.S. and HAUSER, J.R. (1978). Destination choice behaviour for non-grocery shopping trips. *Transportation Research Record*, No. 673, 157-65.

KRISHNAMURTHI, L. and RAJ, S.P. (1988). A model of brand choice and purchase quantity price sensitivities. *Marketing Science*, Vol. 7, No. 1, 1-20.

LEE, L.F. (1983). Generalized econometric models with selectivity. *Econometrica*, 51, 507-12.

MANNERING, F.L. and WINSTON, C. (1985). A dynamic empirical analysis of vehicle ownership and utilization. *Rand Journal of Economics*, 16, 215-36.

McCARTHY, P.S. (1979). Generalized attributes and shopping trip behaviour. *Transportation Research Record*. No. 728, 82-89.

McFADDEN, D. (1981). Econometric models of probabilistic choice. In C.MANSKI and D.McFADDEN (eds.), *Structural Analysis of Discrete Data*, MIT Press, Cambridge, Mass., 198-272.

McFADDEN, D., WINSTON, C. and BOERSCH-SUPAN, A. (1986). Joint estimation of freight transportation systems under non-random sampling. In A. DAUGHERTY (ed.), *Analytic Studies in Transport Economics*, Cambridge University Press, New York.

MOREY, R.C. (1980). Measuring the impact of service levels on retail sales. *Journal of Retailing*, 56, 81-90.

NAG (1978). *Numerical Algorithms Group Library Manuals*, Vols 1-6, Numerical Algorithm Group, Oxford.

NEVIN, J.R. and HOUSTON, M.J. (1980). Image as a component of attraction to intraurban shopping areas. *Journal of Retailing*, 56, 77-93.

PARCELLS, R.J. and KERN, C. (1984). A disaggregate model for predicting shopping area market attractions. *Journal of Retailing*, 60, 65-83.

PODDER, N. (1971). Pattern of household consumption expenditures in Australia. *Economic Record*, 47, 379-398.

RECKER, W.W. and KOSTYNIUK, L.P. (1978). Factors influencing destination choice for the urban grocery shopping trip. *Transportation*, 7, 19-33.

RICHARDS, M.G. and BEN-AKIVA, M.E. (1975). *A Disaggregate Travel Demand Model*. Saxon House, Lexington Books, Farnborough, England

ROY, R. (1942). *De L'Utilite - Contribution a la theories des choix*. Hermann, Paris.

SMALL, K.A. (1982). The scheduling of consumer activities: work trips. *American Economic Review*, 72, 467-79.

STANLEY, T.J. and SEWALL, M.A. (1976). Image inputs into a probabilistic model: predicting retail potential. *Journal of Marketing*, 40, 48-53.

THEIL, J., SUHM, F.E. and MEISNER, J.F. (1981). *International Consumption Comparisons: A System-Wide Approach*. North Holland, Amsterdam, 200pp.

TRAIN, K. (1986). *Qualitative Choice Analysis: Theory, Econometrics and an Application to Automobile Demand*. The MIT Press, Cambridge, Mass.

WALES, T.J. and WOODLAND, A.D. (1977). Estimation of the allocation of time for work, leisure and housework. *Econometrica*, 45, 115-32.

WEISBROD, G.E. PARCELLS, R.J. and KERN, C. (1984). A disaggregate model for predicting shopping area market attraction. *Journal of Retailing*, 60, 65-83.