MODELLING SHOPPING DESTINATION CHOICES:
A THEORETICAL AND EMPIRICAL INVESTIGATION

by

PETER O. BARNARD

A thesis presented for the degree of Doctor of Philosophy in the Department of Economics, University of Adelaide, Adelaide, South Australia

September, 1986

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university and to the best of my knowledge and belief, the thesis contains no material previously published or written by another person, except where due reference is made in the text

PETER O. BARNARD

To my grandmothers Marion Hughes and Doris Marjory Barnard who have lived through times when (of necessity) greater concern was placed on the availability of funds to spend than where such funds might be best spent.

ERRATA

In material like that contained in this thesis it is easy to generate errors. Below is a list of known errors at the time of final submission of the thesis. For drawing my attention to some of these errors I thank the two examiners and a number of colleagues who have reviewed the work subsequent to initial submission. Other of the errors listed have come to light in the process of preparing research papers from the thesis. I apologize to the reader for any remaining errors.

p. xi In line 11 'relationships' should be singular.

p. 32 In line 2 of the footnote 11A should read IIA.

p. 52 Equation (2.24) should read:

$$E_q = f_2(B_{q\ell}, p_{1q}, \ldots, p_{Nq}, b_{1q\ell}, \ldots, b_{Nq\ell}, a_{1q}, \ldots a_{Nq})$$

p. 53 Equation (2.25) should read:

$$T_q = f_3(E_q, b_{1q\ell}, \ldots, b_{Nq\ell})$$

p. 61 Line 10 should read:

$$U_j(g_j^*, \bar{Z}_j^*, L_j^*, B_j, \epsilon_j) > U_i(g_i^*, \bar{Z}_i^*, L_i^*, B_i, \epsilon_i)$$

p. 62 Equation (3.10) should read:

$$g_i = - \frac{\partial V_i / \partial p_i}{\partial V_i / \partial Y} = g_i(p_i, B_i, t_i, c_i, Y, \epsilon)$$

p. 64 In the footnote, line 1, b_i should be replaced with B_i.

p. 69 In line 1 of equation (3.24) p_1 should be replaced with p_i.

p. 77 Equation (3.47) could have been derived by multiplying equation (3.39) by p_j and taking logarithms, rather than the convoluted derivation used in the text.

p. 78 Lines 14/15 should read: 'where v_q is an individual specific error term assumed to be distributed iid extreme value type 1'.

p. 123 Line 8 should read:

$$AVGCONV_d = \sum_q^{Q_d} GCONV_d \; / \; Q_d$$

p. 163 Equation (5.6) should read:

$$E(A^* \mid N) = \int_0^\infty N \, A \, [F(A)]^{N-1} \, dF(A)$$

p. 163 Lines 3/4: the sentence, '$E(A^* \mid N)$ will be a decreasing function of N' should be omitted.

p. 163 Equation (5.7) should read:

$$f(N) = E(A^* \mid N) - E(A^* \mid N-1)$$
$$= \int_0^\infty A \, [F(A)]^{N-1} \Big[N \, [1 - F(A)] + F(A) \Big] \, dF(A)$$

pp. 168/169 A line is missing from the bottom of p.168. The correct wording for the paragraph following equation (5.15) is: ' The model described by equations (5.13) - (5.15) was first constructed (in a different context) by Sheffi (1979). It is one of a family of models that may be applied to ordinal data. Other members of this family are the exploded logit model of Beggs et al. (1981) and the ordered logit model described, for example, in Maddala (1983, pp. 46-49)'.

p. 168 An examiner has astutely observed that the model of equations (5.13) - (5.15) is equivalent to the sequential logit model (see e.g. Maddala 1983, pp. 49-51) jointly estimated and with parameter restrictions imposed. This examiner has also pointed out that the iid property shared by error terms in a logit model is an especially strong assumption in this context.

p. 173 The heading of Table 5.7 should read: 'SIMULTANEOUSLY ESTIMATED SEQUENTIAL LOGIT MODEL OF CHOICE SET SIZE'.

p. 180 Equation (6.1) should read:

$$V_{q(sp)} = \bar{V}_{q(sp)}(Z_{q(sp)}, \alpha) + \epsilon_{q(sp)}$$

p. 201 The footnote should read, 'Note that $\sigma_{u_j\eta_j^*} = \sigma_{u_ju_j}\rho_{u_j\eta_j^*}$ since $(\sigma_{\eta_j^*\eta_j^*})^2$ is equal to 1'.

p. 215 Internal consistency check (iv) should read:

$$\frac{PAR(11)}{PAR(4)} = \frac{\alpha_5\mu}{\alpha_1 a_1^{\alpha_5}} = \frac{PAR(12)}{PAR(5)}$$

pp 215/16. Generally on these pages μ has been omitted from the estimated parameters where an obvious cancellation exists. For example, for consistency check (i) estimated parameters from the MNL model are PAR(3)/μ and PAR(4)/μ. However, $\frac{PAR(3)/\mu}{PAR(4)/\mu}$ has simply been presented as $\frac{PAR(3)}{PAR(4)}$.

p. 217 The last term of the RHS vector should read:

$$\frac{1}{PAR(5)} \alpha_1 a_1^{PAR(12)}$$

p. 219 Line 4 of equation (7.34) should read:

$$- \alpha_2 a_1^{(\alpha_5-1)} p_{iq}^{* a_2(\alpha_5-1)}$$

p. 220 Line 2 of equation (7.39) should read:

$$= V_{iq}\alpha_1^{-1}\alpha_5 p_i^{(\alpha_5-1)} + \alpha_1^{-1}\alpha_2 p_i^{-1}$$

p. 221 Line 4 of equation (7.40) should read:

$$- \alpha_1^{-1}\alpha_2 p_i^{-2}$$

p. 246 Line 3 should read:

$$(\sigma_{u_ju_j})^2 = \frac{1}{Q}\sum (E_{iq} - X_{iq}\beta)^2$$
$$+ (\sigma_{u_j\eta_j^*})^2 \frac{1}{Q}\sum\left[J(\alpha, Z_{iq})\ \phi[J(\alpha, Z_{iq})] \ /\ D(\alpha, Z_{iq})\right]$$

p. 286 The reference with authors listed as 'NAKANISKI, M. and L.F. COOPER' should be 'NAKANISHI, M. and L.F. COOPER'.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This thesis has been materially assisted by a number of individuals and organisations.

In terms of intellectual input, David Hensher provided the main supervisory function for the thesis. I am especially grateful for the time David put into supervising the thesis as this was done in an honorary capacity. His suggestions were invaluable in improving aspects of the work reported herein. My supervisor at the University of Adelaide, Alastair Fischer, carefully read a draft of the thesis, provided encouragement to complete this work, and oversaw the administrative arrangements associated with thesis submission. John Taplin guided the early stages of the thesis.

Sustenance during my time at the University of Adelaide was provided by a South Australian Department of Transport scholarship. The South Australian Department of Transport also part sponsored the survey used for the empirical work in this thesis. I would particularly like to thank the Director General of Transport for South Australia, Derek Scrafton who at all times indicated a willingness to support the project both financially and by more abstract means.

The Australian Road Research Board not only were a part sponsor of the survey used in this thesis but also provided a convivial and stimulating employment environment over the concluding phases of this thesis.

Bob Botterill provided computer assistance in a number of areas, especially in the development of the program of Appendix 7B.

The thesis was typed, in the most part cheerfully, by Adriana Strohmeier, Virginia Young and Marion Donaghey. Special thanks goes to Jan Brownridge who organised the final production of the thesis and who’s views on the importance of this work would seem to exceed my own.

Finally I thank my family and close friends for their support over the years that I have been involved in this study.

ABSTRACT

In this thesis major contributions are made in two areas of study. The contribution made to retail planning is to generalise the Huff model, which has been extensively applied to forecast retail expenditure levels at stores and shopping centres. The contribution made to travel demand analysis is to demonstrate, in the context of shopping travel, that discrete travel choice models, founded on economic random utility theory incorporate to a substantial extent the decisions of travellers regarding activity participation at trip ends. The relationships between travel and activity decisions has been a major area of debate in the transport literature over the past decade.

Both of these contributions arise from the specification of a comprehensive economic theory of shopping destination choice. This theory is so structured to take advantage of findings from mainstream economic consumer theory. An important relationship long ago unearthed in economic consumer theory, Roy's identity, is used to establish a close link between shopping destination choice and retail expenditure.

The empirical counterpart to this theoretical link is an inter-related model of shopping destination and expenditure choices. The relationship between these choices is recognised by, firstly specifying theoretically compatible submodels for the destination and expenditure decisions, and secondly, demarcating the system within the set of sample selectivity models. The estimation of this system is modestly pioneering in a strictly econometric sense.

This thesis also contains a number of minor contributions to the study of shopping destination choice behaviour. Prominent amongst these is a detailed analysis of reported variations amongst individuals in the range of store choices available, with particular reference to the impact these exert on parameter estimates associated with multinomial logit models of shopping travel. Another relatively minor contribution is an analysis of travel linkages between categories of food shopping.

The empirical setting of the study is Adelaide, Australia with the data being derived from a specially conducted survey.

The thesis concludes with some suggestions for further research.

NOTATION

Listed below is the principal set of notation used in this thesis. An attempt has been made to utilize a consistent set of notation throughout. The global set of notation defined below, however, should only be used as a guide. Generally, a detailed definition is provided locally. For instance E_{iq} is defined below as conditional shopping expenditure at destination i by individual q; however, is used in Chapter 7 to refer specifically to expenditure on grocery items. In all cases the local definition overrides the global definition. Much of the local notation is not defined below.

In line with common practice, for each of the X_q, Z_q, G_q and B_i vectors, capital letters have been used to denote the vector and small letters to denote elements in the vector. This, however, does not apply to other vectors. X_q' is used to refer to the transpose of vector X_q, etc. Subscripts delimit bounds for the variable or vector. Thus, for example, given that X_q refers to a row vector of explanatory variables pertaining to individual q included in a continuous choice model, X_{iq} is as X_q but refers especially to alternative i, etc. The q subscript is always used to refer to an individual (q = 1, 2, ..., Q) and the i and j subscripts used to refer to choice (particularly, mode/destination) alternatives (i = 1, 2, ..., N). When i is broken into its components, d is used as the destination subscript and m as the mode subscript. In Chapter 6 sp and $(sp)'$ are used to refer to shopping patterns.

For mathematical operations $\sum$ and Π are conventionally used to signify summation and multiplication and 'log' used to refer to the natural logarithm. The symbol ^ is used with respect to estimated parameter values.

A_d = a vector of variables describing the attractiveness of shopping destination d.

a_1, a_2 = parameters associated with the translation of perceived shopping prices into real shopping prices.

B_d = a vector of quality variables associated with the consumption of shopping goods from destination d ($= b_{d1}$, $b_{d2}, \ldots, b_{dK}$).

C_{qj}^* = the expected retail expenditure by consumer q associated with mode / shopping destination j.

C_m = the set of choice sets containing m alternatives.

C_i = consumption of good i.

c_i = the monetary cost of travel associated with alternative i.

D_{qi} = a vector of variables representing the separation of consumer q from mode / shopping destination i.

D( ) = logit function.

E_{iq} = retail expenditure conditional upon the choice of mode / shopping destination alternative i by individual q ($= p_i q_{iq}^* = (pq^*)_{iq}$).

E(a) = the expected value of a.

E(a|b) = the expected value of a conditional on b.

F_{qi} = a function relating the separation (D_{qi}) and shopping destination attribute variables (A_{qi}) to consumer destination choices.

G = a vector representing consumption of retail goods from shopping destinations, 1, 2, ..., N (= g_1, g_2, ..., g_N).

g_i^* = demand function for consumption of shopping goods conditional upon choice of mode / destination alternative i.

I_q = a polychotomous variable defined for individual q with values 1 to N_q and I_q = j if alternative j is chosen by individual q.

IV = inclusive value.

J_q = the set of objectively available choice sets for individual q.

J( ) = a function which transforms a variable from any well specified distribution to a standard normal variable.

k_{iq} = a binary variable taking value 1 if I_q = i and 0 otherwise.

L = leisure time.

L_i^* = demand for leisure time conditional upon the choice of mode / shopping destination alternative i.

O = a vector of socio-economic, etc. variables.

P_{iq} = the probability that individual q will select alternative i.

$P(j \in N_q)$ = the probability that alternative j is an element in the choice set N_q.

$Prob\{I_q=j\}$ = the probability that individual q will choose alternative j.

p_d = a real price index associated with retail purchases at destination d.

p^*_{dq} = a perceived price index associated with retail purchases at destination d by individual q.

S = a vector of size related destination attractiveness measures.

T = total available time.

t_i = the travel time concomitant with alternative i.

U = direct utility function.

U^* = bivariate direct utility function.

U_i = the conditional direct utility function associated with alternative i.

u = error term associated with a continuous choice model defined with respect to the population at large.

V = indirect utility function.

V^* = bivariate indirect utility function.

V_i = the conditional indirect utility function associated with alternative i.

$\overline{V}_i$ = the 'representative' component of V_i.

v = error term associated with a continuous choice model after allowing for the conditionality of data used for model estimation.

$\tilde{v}$ = an error term associated with a continuous choice model after allowing for data conditionality and the difference between estimated and true selectivity correction factors.

W( ) = a function relating socio-economic characteristics to retail expenditure levels.

X_q = a row vector of explanatory variables associated with a continuous choice model and pertaining to individual q.

Y = income.

Z_{qi} = a super row vector of explanatory variables defined with respect to individual q and alternative i, contained in a discrete choice model and thus associated with the representative conditional indirect utility functions (= A_{qi}, D_{qi}).

$\overline{Z}$ = the Hicksian composite commodity.

$\overline{Z}_i^*$ = demand for the Hicksian composite commodity conditional upon the choice of mode / shopping destination alternative i.

α = a parameter vector of parameters contained in a discrete choice model and thus associated with the representative conditional indirect utility functions $(= \alpha_1, \alpha_2, \ldots, \alpha_R)$.

β = a parameter vector associated with a continuous choice model $(= \beta_1, \beta_2, \ldots, \beta_P)$.

γ_d = a parameter vector associated with the quality index for the d th shopping destination $(= \gamma_{d1}, \gamma_{d2}, \ldots, \gamma_{dK})$.

ϵ = a vector of unobserved influences on utility, $\epsilon = \epsilon_1, \epsilon_2, \ldots, \epsilon_N$, which also form error terms in a discrete choice model.

η_j = a discrete choice model error term associated with the Heckman/Lee selectivity correction method, $\eta_j = \text{Max } V_i - \epsilon_j$ $(i = 1, 2, \ldots, N, i \neq j)$.

η_j^* = η_j transformed into a standard normal variable.

μ = the logistic scale factor.

ξ_i = an indicator function with $\xi_i = 1$ if $g_i > 0$ and $\xi_i = 0$ if $g_i = 0$.

π = pi ≈ 3.1416.

ρ_{ab} = the correlation between random variables a and b.

ρ^2 = McFadden's pseudo - R^2.

σ_{ab} = the covariance between random variables a and b.

$(\sigma_{aa})^2$ = the variance of a.

τ_{ji} = the difference between the representative components of the conditional indirect utility functions associated with alternatives i and j; $\tau_{ji} = \overline{V}(Z_i, \alpha) - \overline{V}(Z_j, \alpha)$.

ϕ = the density function of the standard normal.

Φ = the cumulative distribution function of the standard normal.

ψ_d = the quality index associated with shopping destination d.

ω_{ji} = a discrete choice model error term associated with the Hay/Dubin and McFadden selectivity correction method, $\omega_{ji} = \epsilon_i - \epsilon_j$.

Ω_q = the full objectively determined choice set for individual q.