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 $\sigma_{u_j \eta_j^*} = \sigma_{u_j u_j} \rho_{u_j \eta_j^*}$ 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 $PAR(3)/\mu$ and $PAR(4)/\mu$. 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:

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

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},...., 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.

$\text{Prob}\{I_q=j\}$ = 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.

CHAPTER 1

HISTORICAL AND POLICY SETTING AND STUDY AIMS

1. MOTIVATION FOR STUDYING MODELS OF SHOPPING BEHAVIOUR

The motivation for this thesis is the need to improve the methods used to forecast shopping behaviour. Such forecasts are required in two major areas of planning. In transport planning, forecasts are required of the amount of travel, direction of travel and methods of travel for shopping purposes. In retail planning, forecasts are required of the level of intensity of retail activity.

In addition to consumers, three identifiable groups have a major stake in retail planning; government, developers and retailers. The developer needs to know the best location for a retail development given a number of alternative sites, the optimal size of developments in each location and the optimal mix of each development in terms of ratios of specialty stores, to supermarket space, to department store space, to open area, etc. Retailer operators, to a certain extent, need the same sort of information as developers to facilitate decisions on store location and size. Additionally, operators require predictions on the retail activity effects of store running decisions. These store running decisions can be conveniently grouped into four main areas; those relating to pricing, stocking, service and promotion. These decisions are clearly interrelated. For example, decisions on stocking, service and promotion will impinge on store prices. Furthermore, decisions can be required on quite detailed matters. Some examples are whether to change the number of checkout lanes, aisle widths or brands carried for certain product lines.

The final group with an interest in retail planning is government. Government probably has the most complex set of forecasting needs, as it is charged with the responsibility of assessing the total impact of new shopping centres. Of direct interest to the Government are the costs of servicing the new centre and the extra revenue raised from the centre. In times of high unemployment the Government also has a close interest in employment effects, especially in spatial areas where this is needed. These employment effects have a further indirect positive effect through the economic multiplier concept, but also a possible negative impact on employment in older, existing, shopping areas. Other negative indirect impacts may include environmental effects and indirect transport costs.

In 'Western economies' Government intervention in the construction of new shopping centres has in fact tended to become ever more pronounced over the past 10-20 years. Early laws in most of these economies only required the developer to meet zoning regulations. Intervention was later extended to embrace environmental issues, to require the developer to provide a minimum level of parking, site landscaping and the like. Recently governments have become increasingly concerned with the economic impacts of developments. In the United States this has occurred since the National Environmental Policy Act of 1969. In Australia, a number of planning acts passed in the 1970s gave governments considerable powers to reject developments on the grounds of their economic impact.

In New South Wales, for example, the consideration of economic impact is now broadly interpreted to include the secondary effects of employment (and unemployment) as well as the competitive position between proposed and existing retail operations (Sommerville and Wilmoth 1985). Section 90 (1) of the NSW Environmental Planning and Assessment Act of 1979 states:

> "In determining a development application, a consent authority shall take into consideration such of the following matters as are of relevance to the development the subject of that application:
>
> .
>
> .
>
> .
>
> .
>
> (d) the social effect and economic effect of that development in the locality."

Another Section of this Act, concerning the orderly and economic use and development of land, extends economic impact considerations beyond the immediate locality of the development.

Governments main interest in economic impact would seem to be to preserve the viability of existing commercial areas from competitive intrusion by new shopping centres. Governments would argue that significant private and public resources have already been invested in these centres. Public transport is usually focused around existing centres and the proliferation of car-based shopping centres have a tendency to make public transport even more uneconomic than it is now. It may be parenthetically noted that the effect of a new development on public transport services receives specific mention in the NSW Environmental Planning and Assessment Act of 1979 (Section 90 (1)). The general argument also holds for other services such as roads, water and sewerage works.

It is of interest to note that another Section of the NSW Act, Section 94, sets out a means for local government to levy contributions on developers, in money or in kind, for services and facilities needed to cope with a particular (retail) development. This section of the act has received wide use, especially with regard to the provision of traffic and road works.

Traffic generation estimation lies at the point of convergence between retail and transport planning. Predicting shopping destination choice, however, has also had a history in the traditional transport planning sphere. Typically shopping trips represent between 15-20% of all urban trips. It is important, therefore, to be able to predict accurately shopping travel both in sensitivity analysis of policy measures modifying the operation of an existing transport system and in planning for the transport infrastructure. Reflecting this, models concerning shopping choices normally form one distinct component of urban transport forecasting procedures.

2. A RESUMÉ OF METHODS USED TO PREDICT SHOPPING ACTIVITY

Corresponding to the diversity of groups with a stake in predicting shopping activity is an equally diverse set of forecasting methods. In general, four major methods for predicting levels of shopping activity may be identified; regression based traffic generation analysis, trade-area analysis, gravity models and discrete choice models. A brief review of these methods follows.

2.1 REGRESSION BASED TRAFFIC GENERATION ANALYSIS

The approach favoured by government traffic planning authorities in Australia and the United States, for predicting the impacts of new retail centres, has been to regress shopping centre patronage against a vector of explanatory variables (e.g. Traffic Authority of New South Wales 1980, U.S. Institute of Transportation Engineers 1982). Mathematically expressed, the approach is:

$$PAT_i = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K + e_i \qquad (1.1)$$

where PAT_i is patronage at retail centre i, $x_1, \ldots\ldots, x_K$ are independent variables, $\beta_1, \ldots\ldots, \beta_K$ are parameters and e_i is an error term. Patronage can be segmented by arrival mode, time of day, etc. and a number of models estimated. Once the parameters $\beta_1, \ldots\ldots, \beta_K$ have been estimated using data on existing shopping centres, patronage at a new shopping centre can simply be predicted by inserting the appropriate new centre values for $x_1, \ldots\ldots, x_K$. The set of independent variables is normally restricted to gross leasable area, sometimes broken down into general categories.

The regression based traffic generation approach may be viewed as a conceptually inadequate method for predicting shopping centre patronage. It deals only with the observed outcome of individual decisions, rather than directly addressing the decision process itself. This inadequacy becomes manifest in the set of 'explanatory' variables used under this approach. Generally, patronage at a particular centre may be thought of as dependent on the following factors:

(i) accessibility of consumers to the centre,

(ii) attractiveness of the centre, and

(iii) competition between the centre and other retail centres (where competition is defined by reference to constituent elements of (i) and (ii)).

Of these three factors, only the attractiveness factor is taken into account in regression based traffic generation analysis.

2.2 TRADE-AREA ANALYSIS

Trade-area analysis is the traditional retailing approach to predicting the effects of changes in the retail environment. As applied to construction of a new shopping centre, it involves three primary steps:

(i) survey consumer habits in the existing shopping environment,

(ii) take account of the siting and size of the new centre in relation to existing centres, and

(iii) from these two sources, estimate patronage and turnover for the new centre.

One might argue, in general, that there is little inherently wrong with this process. It, however, has two major flaws in application. Firstly, to apply the process properly requires a new survey for every proposed new centre. Secondly, the process lacks rigour. In the final analysis trade areas are determined by the subjective assessment of the analyst.

2.3 GRAVITY MODELS

In response to these weaknesses with traditional trade area analysis, as early as 1929 there were attempts to develop mathematical models of shopping behaviour. Early attempts utilised laws from Newtonian physics about the interaction of masses. Since these first attempts, there has been a continuous stream of shopping behaviour model development leading to the behaviourally based, statistically efficient and flexible discrete choice models currently in vogue in travel forecasting applications and retail research.

The first forms of mathematical shopping behaviour models were based on a postulate that there should be an interaction between two areas according to the attractive forces of their masses, but interaction would be hindered by the intervening space or distance between the two areas. Expressed simply the relationship is:

$$I_{ij} = \frac{s_i \; s_j}{d_{ij}} \qquad (1.2)$$

where I_{ij} is the interaction between areas i and j, s_i and s_j are the masses (eg. population, floor area, employment) of areas i and j, and d_{ij} is the distance between i and j. Note as the intervening distance increases, the interaction decreases. Also as either of the masses increases, so does the interaction between areas i and j.

Reilly (1929) developed this idea with his law of retail gravitation. Reilly's law states that:

> "Two centres attract trade from intermediate places approximately in direct proportion to the size of the centres and in inverse proportion to the square of the distance from these two centres to the intermediate place."

The mathematical expression for this law is:

$$\frac{PAT_{ik}}{PAT_{jk}} = \left(\frac{s_i}{s_j}\right)\left(\frac{d_{jk}}{d_{ik}}\right)^2 \qquad (1.3)$$

where k denotes the intermediate location, PAT_{ik} represents the number of people from area k drawn to shop in retail centre i, and PAT_{jk} represents the number of people from area k who shop at retail centre j.

The most notable feature of these early attempts to predict retail patronage is the mechanistic nature of the models. Consumers choose shopping centres according to a pre-ordained law established by the analyst. This is obviously unrealistic.

Later formulations of the gravity model hypothesized that shopping choices depended upon vectors of variables measuring the attractiveness of shopping centres (A_1, $A_2, \ldots, A_N$) and vectors of variables representing the separation of consumers from centres, (D_{q1}, D_{q2}, ... D_{qN}). A general version of the gravity model can be specified as:

$$P_{qj} = \frac{F_{qj} (D_{qj}, A_j)}{\sum_i F_{qi} (D_{qi}, A_i)} \tag{1.4}$$

where P_{qj} is the flow of consumers from zone q to shopping centre j, and F_{q1}, F_{q2}, ..., F_{qj}, ..., F_{qN} are functions relating the separation and centre attribute variables to consumer choices. Specific models of the genre shown in equation (1.4) have been used extensively in urban transport planning studies conducted over the last two decades.

Normally the F_{qi}s are specified up to an unknown, fixed, parameter vector. An essential difference between the gravity model of equation (1.4) and earlier attempts is that no longer does the analyst impose an immutable law upon consumers, but rather the mathematical functions for shopping centre choices are in part determined by the revealed preferences of the consumers themselves. Thus the unknown parameter vector is estimated using data on the aggregated choices of individual consumers.

An extension of equation (1.4) that has seen extensive application in the retailing area is to weight centre choice by shopping expenditure. This model, known as the Huff model, can be expressed as:

$$\overline{C}_{qj} = P_{qj} E_q = \frac{F_{qj} (D_{qj}, A_{qj})}{\sum_i F_{qi} (D_{qi}, A_{qi})} W(O_q) \tag{1.5}$$

where, $\overline{C}_{qj}$ = the expected expenditure by consumers living in zone q at shopping centre j, E_q = expenditure on retail goods by consumers living in zone q, O_q = a vector of socio-economic characteristics for consumers living in zone q, and W is a function relating socio-economic characteristics to retail expenditure levels. Equation (1.5) provides a more complete description of levels of retail activity than is possible with equation (1.4).

2.4 DISCRETE CHOICE MODELS (DCMs)

Enhancements to the generalized gravity model have principally involved interpreting equations (1.4) and (1.5) at an individual level. Thus P_{qj} becomes the probability that individual q will select shopping centre j and C_{qj}, the expected expenditure by individual q at shopping centre j. Statistical methods for estimating equation (1.4) using individual level data were developed during the late 1960s.

Interpretation of these equations at the level of the individual proffers advantages in two areas. Firstly, estimation at an individual level tends to avoid statistical problems, such as ecological fallacy, associated with the use of aggregated data. Several studies (e.g. Fleet and Robinson 1968, McFadden and Reid 1975) have shown that aggregation of data to the transport study zonal level can eliminate more than half the total data variability. Also, in practice, estimation of these models at an individual level has tended to promote the introduction of a richer range of independent variables. Whereas typical gravity model applications used only distance or travel time as a measure of separation and shopping centre size as a measure of attractiveness, discrete choice model applications have usually specified D_{qi} to include travel time and travel cost and have often used attitudinal measures in the A_{qi} vectors. Secondly, a rationale for the models can be directly obtained from economic consumer theory (McFadden 1975, Domencich and McFadden 1975).

It can be seen that simple manipulation of equations (1.4) and (1.5) provides many predictions of interest to the retailer or transport planner. For instance, summing over q for a particular shopping destination j gives the expected number of people visiting that destination. Similar calculations can be made for shopping expenditure. For the most common specification of F_{qi}:

$$F_{qi} = \exp(Z_{qi}\,\alpha)$$

where Z_{qi} contains all variables in vectors A_{qi} and D_{qi} and α is a vector of parameters, direct travel and attractiveness elasticities can easily be obtained by noting:

$$\frac{\partial \log P_{iq}}{\partial \log z_{iqk}} = \alpha_k\, z_{qik}\,(1 - P_{iq})$$

and cross elasticities by:

$$\frac{\partial \log P_{iq}}{\partial \log z_{jqk}} = -\,\alpha_k\, z_{iqk}\, P_{jq}$$

The past 5 years have witnessed a growing acceptance of the role of discrete choice models in forecasting shopping decisions, particularly shopping destination choices. Developed originally in the transport planning area, discrete choice models concerning shopping behaviour have recently infiltrated the retail planning literature (e.g. Arnold et at. 1983, Eagle 1984, Weisbrod et al. 1984) and there have even been suggestions that this model genre could be useful in traffic generation work (Barnard and Brindle 1985, Verster et al. 1979). Despite increasing acceptance, however, discrete choice models of shopping behaviour may be regarded as still in infancy, with many issues of an empirical and theoretical nature as yet unresolved. It is the aim of this thesis to contribute to a resolution of some of these issues.

3. AN OUTLINE OF THIS THESIS

In particular, this thesis concentrates on three issues in estimating discrete models of shopping destination choice. After a fairly extensive review of existing literature (Chapter 2), a theoretical model of shopping destination choice is constructed (Chapter 3). To the author's knowledge this is the first time discrete shopping destination choice models have firmly been grafted to the roots of economic theory. This theory has two aspects. One is the discrete choice of shopping destination. The other is the continuous choice of shopping expenditure. The theory indicates how these two choices are intricately linked.

The data set used throughout this thesis is described in Chapter 4 and some basic models of food shopping destination choice are estimated. The following two chapters explore some issues in the estimation of discrete shopping destination choice models. In Chapter 5 the specification of choice sets in these models is investigated. In Chapter 6 linkages between various facets of food shopping destination choice are examined. Chapter 7 involves the estimation of an integrated model of grocery shopping store choice and shopping expenditure. The concluding chapter highlights important results from this study and suggests areas for further research.

The binding agent of this thesis is the theoretical exposition in Chapter 3. A number of aspects of this theory are empirically explored in remaining chapters. No attempt is made here to combine these empirical findings into a 'super' model of shopping destination choice. This is too large a task for the purposes in hand. Rather it is the opinion of the author that at this stage it is more important to gain a deeper understanding of the processes inherent in individual shopping destination choices than attempt to build an all encompassing model (see also Hanson 1979).

In this vein, the major contribution of this thesis is to show how economic theory can be used to understand shopping destination choices and the link between these choices and shopping expenditure decisions. This link is also empirically established in the thesis. Minor contributions involve demonstrating systematic variations in reported shopping destination choice sets and the impact of this on model estimation, and how destination choices for the various categories of food shopping may be considered under a unifying model framework.

CHAPTER 2

MODELS OF SHOPPING DESTINATION CHOICE: A REVIEW CONCENTRATING ON MULTINOMIAL LOGIT FORMS

1. INTRODUCTION

The past decade and a half has seen significant changes occur in the techniques used to forecast shopping destination choices. This period has seen increasing acceptance in the transport and retailing communities of probabalistic choice models estimated at the individual or household level in preference to the aggregate and sometimes deterministic structures used in the 1950's and 1960's. Although these choice models have in turn come in for criticism, the fact remains that they have the advantage of being based on an explicit theory of choice behaviour, are data efficient and offer much flexibility in modelling structures.

It is the purpose of this chapter to review this new generation of shopping destination choice models. While the emphasis of this review is on model use in transport planning, partly because this has been the main area of development and partly because it represents the author's primary research interest, the retailing literature has not been neglected. Slightly more stress is placed on food shopping in this review than other forms of shopping. Food shopping was chosen for emphasis as it represents the most frequent repetitive urban trip with a significant short term spatial choice component, accounting for approximately 10 per cent of total person trips (Barnard 1982, Burnett 1974). It is also the area of concentration in the empirical chapters of this thesis. However, most transport related shopping destination choice modelling has been conducted at a more aggregate level, ignoring the type of shopping. Thus, while a focus of this review is food shopping, other types of shopping are not neglected.

By restricting this review to discrete choice models of spatial shopping behaviour, based on random utility theory, many other model forms have been ignored. These include gravity models, entropy models and some attitudinal model structures. The review is not as narrow as may first appear, however, since it can be shown that the singly constrained gravity model is a special case of the logit formulation of discrete choice models (Cochrane 1975, Daly 1982). Logit and gravity models have dominated applied transport and retailing work in modelling spatial shopping behaviour.

The remainder of this chapter is divided into five substantive sections. In Section 2, I discuss modelling objectives and areas of application for spatial models of shopping behaviour. Section 3 contains an outline of the logit formulation of a discrete shopping destination choice model. Highlighted in this section is the identification of attributes to be included in these models. The next three sections examine some further issues in model estimation. The issues examined are choice set definition (Section 4), quasi-dynamic and multi-destination choice models (Section 5) and linking destination choice models with activity measures (Section 6). Important topics not covered include aggregation and model transferability.

2. SHOPPING DESTINATION CHOICE MODELLING OBJECTIVES

Modelling of shopping destination choice, in contrast to the modelling of mode choice, has been characterised by a multitude of mathematical and theoretical structures. Moreover, even within a particular methodology, such as discrete choice modelling, there remains a diversity of approaches in application. Given the complexity of shopping destination choice this wide diversity in modelling approaches is hardly surprising and need not be indicative of a piecemeal approach to the subject. The large variety of temporal, spatial and personal factors which intermix to produce destination choice may make it impossible to capture the essence of choice under all circumstances. The wisdom in choosing a particular approach, then, can only by judged by reference to the specific purpose of the model in question.

Unfortunately, objectives in modelling have often not been made explicit. There appears, however, to be an underlying assumption in the majority of research into travel and retailing choices that the sole purpose of modelling is for prediction. As noted by Stopher and Meyburg (1975) among others, it is clear that prediction in itself incorporates many objectives and is but a subset of uses to which models are put. A possible classification of modelling objectives is:

1. Understanding

   (i) as a general guide to policy
   (ii) for further model development

2. Prediction

   (i) long term forecasts
   (ii) short term forecasts

This scheme divides modelling objectives into two general classes : prediction and understanding. Given the incomplete nature of choice models of all forms, coupled with Heggie's (1977) observation that an astute policy maker can mostly predict behavioural changes better than a model, understanding is possibly the more important of these objectives. Stopher (1979) in support of this proposition affirms that understanding 'may have more policy impact than any other role or capability'. An example provided by

Stopher is the effect that the development of models postulating a choice of modes between private and public transport - in contrast to the mode split models of the 1950's, which were based almost exclusively on socio-economic descriptors of zones - had on influencing policies (either by way of initiation or confirmation) aimed at attracting people from car to public transport. Models can also be built simply for research purposes as an aid to further model development. Some attitudinal models may fall into this category which, while not of direct policy relevance, may nevertheless draw attention to salient attributes of a particular decision for further research to derive appropriate policy manipulatable measures. Presumably the substantial number of research studies, appearing in the transport literature, which develop models containing perceptual measures of shopping destination attractiveness, fall into this category (e.g. Koppelman and Hauser 1978, McCarthy 1979, Recker and Kostynuik 1978, Stopher 1979, Timmermans et al. 1982).

It is also possible to divide the prediction objective into short term and long term forecasts and each of these will have different modelling needs. Long term forecasts are required to support the planning of the transport infrastructure over a period of ten or more years. Predictions of this type require that all independent variables be readily forecastable and an allowance be made in the modelling structure for interaction between transport and other sectors. For these models it is usually necessary to define shopping destination alternatives as fairly large geographical areas (zones).

Short term shopping destination choice predictions may be useful in two areas of transport decision-making. One is in planning the transport infrastructure, especially the supply of parking, necessary to support a new store, shopping or activity centre. This application of discrete shopping travel choice models has received little attention in the literature. However, Barnard and Brindle (1985) note several advantages of these models over the traffic generation models that are currently in use. The essential difference between models developed for this purpose and those used for longer term forecasting, is that the shorter prediction period makes it possible to introduce a richer set of independent variables, such as the socio-economic characteristics of individuals having the new centre as a shopping alternative. Also these alternatives can now be defined on an individual shopping centre, rather than zonal, basis. However, destination attractiveness measures must be restricted to those that are known to the planner at the time of the development application. This effectively reduces the set of available measures to parking supply and floor space.

The other main area in transport for application of destination choice models is in sensitivity analysis of policy measures modifying the operation of an existing transport system. A particular application is examining the effect on the distribution of shopping trips of a change affecting use of one or more transport modes (e.g. an increase in petrol prices). In these circumstances better travel forecasts may well be obtained by adopting attitudinal measures for variables (such as destination attractiveness descriptors) that in any case would be outside the control of a transport policy maker, thereby resulting in a more correctly specified model and increasing the accuracy of parameter estimates associated with the transport related variables that are amenable to manipulation (Richards, 1979).

Shopping destination choice models are also very useful in retail planning. Typically destinations are identified as individual stores. In addition to travel variables, detailed destination descriptors can be included in these models. Examples are indices of store prices, selection, quality of merchandise, number of checkout lanes, etc. As such, retail models normally contain a higher level of behavioural content than those used in transport.

In view of this discussion, in the review that follows, particular attention is paid to the circumstances in which the model may apply. This is especially evident in the next section, with its discussion on measures of destination attractiveness.

3. BASIC MNL MODELS OF SHOPPING DESTINATION CHOICE : AN OVERVIEW

The model formulation that is subjected to concentrated attention during this review is the multinomial logit model. It has the general form:

$$\text{Prob}\ \{I_q=j\} = P_{qj} = \exp \overline{V}_{qj} \Big/ \sum_{i \in N_q} \exp \overline{V}_{qi} \qquad (2.1)$$

where I_q is a polychotomous variable taking values 1 to N and I_q = j if alternative j is chosen by individual q, N_q represents the set of alternatives available to individual q, and the $\overline{V}_{qi}$ are parameterisable functions consisting of variables describing the attractiveness of alternative i (i = 1, 2, .., j, .., N_q) to individual q. $\overline{V}_{qi}$ can be interpreted as an index of utility obtained by individual q when consuming alternative i. In particular, an MNL model can be derived by assuming that the total utility, V_{qi}, obtained by individual q when consuming alternative i consists of an identifiable representative component, $\overline{V}_{qi}$, and an unobservable component, ϵ_{qi}, where the latter are independently and identically distributed (iid) extreme value type 1 across individuals and alternatives (McFadden 1974):

$$V_{qi} = \overline{V}_{qi} + \epsilon_{qi} \qquad (2.2)$$

Alternatives in the context of spatial choice shopping modelling can be defined as shopping destinations (d) or mode (m)/destination combinations. Shopping destinations can be identified as zones, shopping centres or individual stores.

$\overline{V}_{qi}$ can be conveniently decomposed into two variables, one representing the disutility of travel to shopping destination i, D_{qi}, and the other representing the attractiveness of the shopping destination, A_{qd}:

$$\overline{V}_i = \overline{V}_{qi}(D_{qi}, A_{qd}) \tag{2.3}$$

In turn each of these variables can be endogenised by expressing D_{qi} as a function of fixed parameters, α_i^1, and characteristics of the transport system serving i, Z_{qi}^1, and A_{qd} as a function of fixed parameters, α_i^2 , and destination attractiveness descriptors, Z_{qd}^2.

There seems to be general agreement that the Z_{qi}^1 included in D_{qi} can be adequately represented by travel times and costs to destination i. When the travel alternatives used in the logit model are taken to be mode/ destination combinations, the modal specific travel times and costs can be included directly in the model. However, when travel alternatives are defined exclusively as destinations, a composite accessibility measure must be constructed, which includes elements of travel by all available modes. Many ad hoc accessibility measures have been used. Following work by Ben-Akiva and Lerman (1979), however, it is now recognised that the theoretically correct measure, when the shopping destination choice model specified is of the MNL form, derived from random utility theory, is defined by an inclusive value term calculated from a lower level model explaining choice of mode. From such a model levels of utility, $\overline{V}_{qdm}$, associated with each modal alternative, m, available to destination d can be identified and the inclusive value index, IV_{qd}, calculated as:

$$IV_{qd} = \log\left[\sum_{m}^{M_{qd}} \exp(\overline{V}_{qdm})\right] \tag{2.4}$$

where M_{qd} is the set of modes available to destination d for individual q.

There is much less consensus on the terms to be included in the vector Z^2_{qd}. Contained in Table 2.1 is a listing of major studies concerning shopping destination choice that have appeared in the transport literature in the last decade. Also included are a selection of studies reported in the marketing literature that have involved the estimation of discrete destination choice shopping models*. Basically two groups of studies can be recognised.

In one group, studies are characterised by the use of aggregate (zonal) concepts of shopping destinations and non-behavioural, somewhat circular, measures of destinational attractiveness, such as retail employment and selling space. Heggie (1977) has argued that theoretically such measures are of dubious validity. It is more probable that retail employment and provision of selling space are a consequence of the spatial pattern of demand, rather than measuring destination attraction. There is a certain tautology implied in the inclusion of variables, themselves a function of aggregate travel choices in models purporting to predict these choices. One study (Southworth 1981) simply used the absolute level of zonal trip attractions, estimated from a category model, as the measure of destination attractiveness. This practice has been common when destination choice models are estimated as part of a transport planning exercise (e.g. Pak Poy and Assoc. 1978). For consistency, when such measures have been adopted, an iterative process should be invoked in which numbers of trips terminating in each zone, as estimated from the destination choice model are checked against those

---

* A notable omission in the coverage of marketing research, is reference to any studies concerned with where to purchase particular products or brands. This has been a major use of discrete choice models in marketing. The lowest level of aggregation considered in this review is where to purchase classes of products (e.g. groceries).

TABLE 2.1

A LISTING OF MAJOR EMPIRICAL STUDIES CONCERNED WITH MNL MODELLING OF SHOPPING DESTINATION CHOICE

STUDY REFERENCE	DATA SOURCE (SAMPLE SIZE)	PURPOSE CATEGORY(1)	IDENTIFICATION OF CHOICE ALTERNATIVES	DEFINITION OF DESTINATION	METHOD USED TO IDENTIFY DESTINATION ATTRIBUTES	DESTINATION ATTRIBUTES USED
GROUP I						
Adler and Ben-Akiva, 1976	Washington, USA, 1978 (403)	all shopping	frequency/ destination/ mode	zones	researcher specified	retail employment, CBD dummy
Barton-Aschman Assoc/Cambridge Systematics, 1976(2)	Los Angeles, USA	noon hour CBD shopping	frequency/ destination/ mode	zones	researcher specified	retail floor space ÷ zonal area, zonal area
Ben-Akiva et al., 1980	Washington, USA, 1968	all shopping	mode/ destination	zones	researcher specified	retail employment, retail employment ÷ net commercial acreage
Charles River Assoc, 1976 (Domencich and McFadden, 1975)	Pittsburgh, USA, 1977 (63)	all shopping	destination only	zones	researcher specified	retail employment
Kitamura and Kermanshah, 1984	Baltimore, USA, 1977 (130)	all non-home based shopping	destination only	zones	researcher specified	zonal population, composite retail employment, store hours variable
Landau et al., 1982	Tel Aviv, Israel (467)	all shopping	mode/ destination	zones	researcher specified	retail employment
Miller and O'Kelly, 1982(2)	Ontario, Canada, 1978 (356)	grocery shopping	destination only	zones	researcher specified	retail area
Richards and Ben-Akiva, 1975	Eindhoven, The Netherlands, 1970 (430)	all shopping	mode/ destination	zones	researcher specified	retail employment, CBD dummy
Ruiter and Ben-Akiva, 1978	San Fransisco, USA, 1965	all shopping	mode/ destination	zones	researcher specified	retail employment, retail employment ÷ population serving acreage, CBD dummy
Salmon and Ben-Akiva, 1983	Baltimore, USA, 1977 (344)	all shopping	mode/ destination	zones	researcher specified	retail employment, retail employment ÷ total employment, CBD dummy
Southworth, 1981	West Yorkshire, England, 1970 (517)	all shopping	mode/ destination	census zones	researcher specified	number of zonal trip destinations
Swait et al., 1984	Maceio, Brazil, 1977 (234)	all shopping	frequency/ destination/ mode	zones	researcher specified	employment (service and professional), CBD dummy, CEASA market place dummy
Van der Hoorn and Vogelar, 1978	Amsterdam, The Netherlands, 1976 (795)	all shopping	destination only	zones	researcher specified	employment (tertiary)

GROUP 2

Arnold et al., 1983	various sets	food shopping	destination only	stores	researcher specified	prices, variety, service, quality, weekly specials, value, fastest checkout counters, pleasant shopping environment (all 0-1 variables)
Eagle, 1984	experimental design, unspecified location	grocery shopping	destination only	synthetic stores	factor listing	price, selection
Gautschi, 1981[3]	San Fransisco, USA, 1973/74 (350)	major non-grocery shopping	mode/ destination	shopping centres	researcher specified + factor analysis	assortment, centre design, low prices, hours, dress, lack of crowds[4]
Ghosh, 1984[3]	location unspecified (300)	major grocery shopping	destination only	stores	researcher specified	store size
Koppleman and Hauser, 1978 (Stopher, 1979)	Chicago, USA ,1974 (500)	non-grocery shopping	destination only	shopping centres	researcher specified + factor analysis + MDS	variety, quality and satisfaction, value, parking[4]
Recker and Kostynuik, 1978	Buffalo, USA (300)	major shopping	destination only	stores	researcher specified + factor analysis	quality, store convenience, service, store type
Timmermans et al., 1984	Eindhoven, The Netherlands, (91)	clothes shopping	destination only	shopping centres	repetory grid	number of stores, parking
Verster et al., 1979	Amsterdam, The Netherlands, 1974/75 (200)	various shopping purposes	mode/ destination	shopping centres	researcher specified	retail floorspace by category, consumer perceived prices, researcher perceived prices
Weisbrod et al., 1984	Boston, USA, 1977 (170)	major shopping	mode/ destination	shopping centres	researcher specified	total number of stores, proportion of clothing and general merchandise stores, proportion of other stores, variety store, planned centre dummy.

Notes: (1) Categories refer to home-based shopping unless otherwise specified
(2) Minor studies included because of special features
(3) Model estimated using frequency data
(4) Refer to factor dimensions.

obtained from the trip attraction model, adjusted if necessary, the models re-estimated, and the process continued until coincidence of estimates is achieved.

Many studies in this first group have also included a dummy variable taking values of 1 if the destination zone lies within the CBD and 0 otherwise. This variable, in effect, captures qualitative differences between the CBD and suburban shopping destinations. A more general way to capture these differences is to employ a nested logit model (Sobel, 1980). An intuitively appealing nested logit schema is shown in Figure 2.1. The lower-level model in this schema concerns the choice between alternate suburban shopping centres. From this model an inclusive value index can be constructed, encapsulating the net attractiveness of all suburban shopping centres, and used as input into a binary logit model of the choice between the CBD shopping centre and suburban shopping centres (treated as a group). It can be shown that the nested logit model is a generalisation of the simple logit model of equation (2.1) (McFadden 1978). It can be estimated either with a two-stage process (e.g. Ortuzar and Donoso 1983) or using full information maximum likelihood (e.g. Hensher 1986). Nested logit has yet to be extensively applied in a destination choice context.

The second group of studies in Table 2.1 have used a variety of psychometric methodologies in an endeavour to identify behavioural measures of destination attractiveness. A popular approach has been to present shoppers with a list of specified attributes, typically of the order of 10-20 attributes, derived from literature reviews and/or questionnaire pretesting. Respondents are then required to rate possible shopping destinations by each attribute using scales such as the Likert or semantic differential scales or by paired comparisons. A danger with some of these questioning methods is that biases may be introduced due to inclusion of irrelevant attributes or exclusion of relevant attributes (Timmermans et al. 1982).

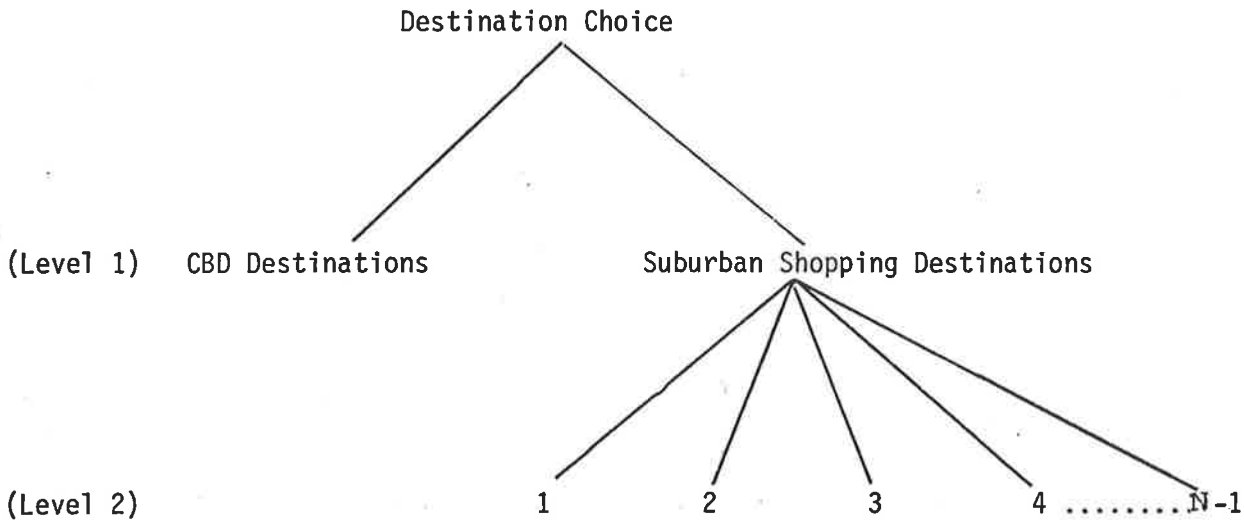


FIGURE 2.1: A SUGGESTED NESTED MODELLING STRUCTURE FOR SHOPPING DESTINATION CHOICE

Either attribute scores can be included in a destination selection model in their raw form or a data reduction technique applied such as factor analyses or multi-dimensional scaling (MDS). A problem with including raw attribute scores is that results may be unstable due to multicollinearity. Data analysis techniques such as factor analysis serve to remove multicollinearity but since these techniques are purely statistical, relying on correlation between attribute scores, the resultant factor loadings are often difficult to interpret in a behavioural dimension. For this reason it is difficult to justify the interpretation of factors as identifying underlying cognitive dimensions (see, for example, Koppelman 1980, p. 140 and McCarthy 1979). Association in the mind is just one explanation of correlated attribute scores. Attributes independently cognised may nevertheless be correlated because of quite unrelated considerations. As an example, economies of scale may mean that attribute scores for prices and variety will be correlated and hence loaded onto one factor even though they are perceived by the respondent on dissimilar dimensions.

Another strand of research designed to uncover behavioural measures of destination attractiveness is by factor listing. Respondents are simply asked to specify the reasons for preferring their chosen destination to other available destinations. Responses are then content analysed and subjected to frequency of mention counts. Normally it is found that most respondents specify a common small set of attributes (2-4) as entering their decision-making process. These attributes can then be combined into an experimental design (e.g. Louviere and Meyer 1981) or destinations rated by each attribute and incorporated directly into a destination choice model (e.g. see Chapter 4). This method of identifying the cognitive dimensions of destination attractiveness avoids the problems associated with factor analytic methods, is less complex in design, and the dimensions derived in this way have been shown to correlate well both with observed behaviour and with physically measurable destination attributes.

A more formalised method of obtaining respondent specified destination attractiveness measures is by Kelly's repertory grid. This technique commonly involves presenting respondents with triads of elements (stores, shopping centres) and requesting that they specify some important way in which two elements are alike but different from the third. Different triads continue to be presented to each respondent until no new constructs are specified. On completing this phase respondents are required to rate each element on each construct he/she has provided. Finally, ratings of importance of each construct to the individual are obtained. The repertory grid method has been used by Timmermans et al. (1982) in a study of choice of shopping centre for daily and non-daily goods. Results were compared to the factor listing approach and found to be very similar. In view of time consuming interviewing process associated with the repertory grid method it is likely that factor listing will be preferred for many applications.

As can be appreciated from the foregoing discussion each method designed to uncover behavioural dimensions of destination attractiveness has associated weaknesses. What is impressive, however, is that very different methods have revealed very similar cognitive dimensions of destination attractiveness. For destinations defined as stores, prices, variety, quality of merchandise, and parking facilities have dominated study results, irrespective of method. For shopping centres the range and number of stores and centre atmosphere are additional factors*. Unfortunately, although it is relatively easy to derive direct physical measures for most of these dimensions (e.g. real price indices for perceived prices) it is difficult to predict these measures other than over a very short time period. It is difficult

---

* Of course, in virtually all studies accessibility factors have also been found of importance.

to envisage, for instance, how indices representing prices, variety and quality of goods, can be used in traffic generation work or in medium to long term transport planning exercises. On the other hand the inadequacy of currently used measures of destination attractiveness, such as retail employment or selling space, has been previously noted. What is needed is research aimed at interfacing these two divergent approaches. There has been an appalling lack of research aimed at relating behavioural measures of destination attractiveness to readily predictable and stable variables.

Not only is the identification of appropriate destination attractiveness measures a critical issue, but also the form in which they are included in the model. This especially applies to size related measures of attractiveness, such as employment or selling space. In particular, it is desirable that the choice probabilities be directly proportional to these variables. Also when destinations are defined as zones it is desirable that travel forecasts be invariant to any aggregation or disaggregation of zones. For instance, if some change is made to an essentially arbitrary zoning system in which two zones are amalgamated into a single zone, the total number of trips forecast for the new zone should be the summed number of trips for the original zones. Watanatada and Ben-Akiva (1978) and Kitamura et al. (1979) have shown that an appropriate method of introducing size related attractiveness variables is:

$$\overline{V}_{qi} = \overline{\overline{V}}_{qi} + \log\left(\sum_{\ell=L+1}^{M} \alpha_\ell \; s_{qd\ell}\right) \qquad (2.5)$$

where $\overline{\overline{V}}_{qi}$ contains all terms in $\overline{V}_{qi}$, except the size related destination attractiveness measures, s_1, s_2, .., s_L. By combining (2.1) and (2.5) it can be seen that the choice probabilities are directly proportional to the variables s_1,

s_2, .., s_L. It should be recognised that the parameters α_{L+1}, α_{L+2}, .., α_M are only identifiable up to a multiplicative constant and, for this reason, it is necessary to impose a normalisation such as $\alpha_M = 1$. This method of including size variables is now regularly followed in advanced studies belonging to group 1. What has yet to be recognised is that the proportionality argument applies with no less force to variables such as 'number of stores in a shopping centre' or even perhaps to selection and variety of goods within a store*.

4. DESTINATION CHOICE SET IDENTIFICATION

A necessary input in estimating any discrete choice model is specification of a set of alternatives, termed the choice set, from which selection is made. For the model to be behaviourally based, embedded in the constructs of random utility theory, the choice set used by the analyst in the estimation process should correspond to (or represent a subset of) the set of alternatives actually considered by the individual when making a choice. An important feature of the MNL formulation of equation (2.1) is the ability to allow the availability of alternatives to vary between individuals. This ability is highlighted by subscripting the set of travel alternatives, N, from which a choice is made, by q. This formulation allows for the possibility that different individuals may face entirely different choice sets; some individuals may be very restricted in their choice of shopping destinations, while others may have a wide range of possible options. Three general issues are required to be addressed when specifying choice sets.

---

* The latter suggestion follows from an hypothesis that the elemental alternatives are not shopping centres or stores, but products. In the marketing literature it has been common to estimate logit models at this level, with alternatives being defined in terms of spatial locations available for purchasing a particular product.

The first is the overall dimensions of behaviour to be incorporated in model structure. This decision should be made having regard to the policy areas of model use and the type and degree of behavioural response expected for each variation in policy (Heggie and Jones 1978). Two behavioural dimensions have already received prominence in this review : destination choice and mode choice. However, if the policy maker is interested in the degree of travel to each destination under alternate scenarios, a trip generation dimension is obviously also important. Another example is for traffic generation work where 'duration of stay' will affect policy decisions on parking. One recent thrust of research into destination choice models has been to expand the dimensions of behaviour considered and introduce more complex interactions between the various dimensions. In some cases this has involved modelling entire travel patterns, with past destination and activity choices impacting on present and future choices. These models are reviewed in the next section.

A second decision concerns segmenting the dimensions of behaviour into submodels within the overall structure. The principal consideration bearing on this decision for the MNL model is possible violation of the independence from irrelevant alternatives (IIA) property*. It is more likely that the effects of omitted variables, measurement error, etc. will cause the unobserved components of utility to be intercorrelated within a behavioural dimension than between dimensions. When, for example, behavioural dimensions are combined to form alternatives, as in the estimation of a joint mode/destination choice model, it is possible that modes to a particular destination will be similar in their unobserved characteristics. This will result in violation

---

* The IIA axiom states that the ratio of the probabilities of choosing one alternative over another is unaffected by the presence or absence of any additional alternatives in the choice set. It is an inherent property of the MNL model.

of an assumption used to derive the basic logit model, that the correlation between the unobserved attributes associated with each and every pair of alternatives is zero. A solution is to estimate, and appropriately link, separate logit submodels for each behavioural dimension. As was evident from discussion in the previous section, there are also circumstances where it is expedient to subdivide a behavioural dimension for modelling purposes.

The final aspect of choice set specification that requires consideration is the identification of the set of alternatives within a behavioural dimension from which a choice is made. It is known that incorrect inclusion of an alternative in the choice set will result in biased model parameter estimates and understated probability values for the other alternatives. The spatial context of destination choice behaviour poses special problems in appropriately defining a set of alternatives (Ansah 1977, Black 1984, Tardiff 1979). Work on this issue is reviewed below.

The most common approaches to defining destination choice sets have been either to consider only a spatially concentrated portion of the population so as to limit the feasible choices or to prespecify the destination outlets to be included, or both. For example, many transport studies have attributed as destination alternatives the chosen destinations of other individuals living in the same zone (e.g. Project Bureau for Integrated Traffic and Transport Studies 1977, Charles River Associates 1976, Richards and Ben-Akiva 1975). Others have used more loosely defined choice sets such as regarding all destinations in a metropolitan region as being potentially available (e.g. Pak Poy and Associates 1978). Also interesting variations on these general criteria are apparent. As an example, Adler and Ben-Akiva (1976) in their study of shopping destination choice included additional destinations 'based on deductive notions of the perception of alternatives'. The final approach, especially, limits the potential transferability of the model, and all approaches render the models susceptible to problems arising from inclusion of irrelevant alternatives*.

Ansah (1977) was one of the first researchers to define the destination choice set in a more systematic manner. He simplified the choice set generation process by initially assigning physical destinations into functional destination categories. Functional classifications were defined for each origin and based on prespecified ranges of accessibility, represented by airline distance, and attractiveness, measured by retail floor area. The classifications were also separately developed for various market segments. A multinomial response model was used to predict the probability of selecting a particular functional destination category. Although this represented the extent of Ansah's empirical work he further discussed how a Monte Carlo procedure might be invoked to assign functional categories to each origin and a MNL model then used to predict the probability of selecting a particular physical destination from the functional category.

A tangential approach to that of Ansah has been developed by Black (1984) who combines a destination classification scheme with the employment of the threshold values that, if exceeded, for key variables, exclude consideration of a destination or class of destinations. The crucial feature of Black's method is the setting of the threshold values. This task is assisted by construction of a simple index which endeavours to capture the benefits and losses to the modeller of excluding destinations from the choice set. Losses arise from two sources: exclusion of a

---

* This phrase is often used, in this thesis and elsewhere, when discussing a particular property, the IIA property, of MNL models. From the previously provided definition, a more useful mnemonic for the IIA property might be ERA - exclusion of relevant alternatives. Here the phrase 'inclusion of irrelevant alternatives' is used to refer to the inclusion, in the estimation choice set, of alternatives not actually considered by the individual. Problems that may arise from inclusion of such alternatives are explored in detail in Chapter 5.

relevant alternative from the choice set or inclusion of an irrelevant alternative*. Benefits derive from improvements in predictive accuracy by excluding irrelevant alternatives from the choice set. In his empirical work Black compared predictions from probabilistic choice models assuming unconstrained destination choice sets with choice sets filtered by destination specific threshold values. Prediction errors were found to be smaller when choices sets were restricted than for the unconstrained case. This result, in fact, almost definitionally follows from the method used by Black to set the threshold values. However, the size of the gain in predictive efficiency may surprise those who argue that choice set identification is unimportant because probabilities assigned by a 'well specified' model to destinations incorrectly included in the choice set will be so small as to be, in practice, indistinguishable from their true zero value.

The efforts of Ansah and Black, although contributing to knowledge of spatial choice set generation processes, may be criticised in terms of lacking a sound behavioural base. A particular weakness is that the classification scheme and threshold values depend only on destination attributes, including the location of the destination relative to defined origin points. Both the schemes proposed by Ansah and Black cannot readily take into account characteristics of the individual, other than by the data intensive method of market segmentation. Black (p. 67) in defending this aspect of his approach asserts that individual characteristics are unimportant. Evidence provided in Chapter 5 calls into question this assumption.

One approach to increasing the behavioural content in choice set definition has been taken by Burnett and Hanson (1982). They conceived that the entire choice process could be effectively modelled through a decomposition:

---

* Exclusion of a relevant alternative while not affecting parameter estimation in share models with the 11A property, will affect predictive use of such a model.

$$P_{qj} = P(j \in N_q)\ P(I_q = j \mid j \in N_q) \qquad (2.6)$$

The first term on the RHS of equation (2.6) is the probability that alternative j is in the choice set of the individual and the second term refers to the probability of selecting j given that j is contained in N_q.

A more general formulation, firmly rooted in economic random utility theory, has been developed by Manski (1977) and extended by Williams and Ortuzar (1982). In Manski's formulation the choice of an alternative, say j, is represented by two probabilities; the probability, $P(N_q|J_q)$, that an individual selects for consideration a particular subset of alternatives, N_q, from the set of all objectively available choice sets, $J_q(J_q = 1, 2, \ldots, M_q, \ldots, N_q, \ldots, \Omega_{Mq})$, and then, given the choice set, the probability, $P(I_q = j|N_q)$, that alternative j is selected. To obtain the unconditional probability that alternative j is selected, is it necessary to sum the conditional probabilities over all choice sets that are elements of J_q:

$$P_{qj} = \sum_{C \in J_q} P_q(C|J_q)\ P(I_q = j|C) \qquad (2.7)$$

Williams and Ortuzar (1982) further decomposed the probability of selecting N_q from J_q, $P(N_q|J_q)$, by grouping choice sets within J_q that contain equal numbers of alternatives. An expanded version of equation (2.7) can then be written:

$$P_{qj} = \sum_{m=1}^{M_q} \sum_{C \in C_m} P_q(C_m|J_q)\ P_q(C|C_m)\ P(I_q = j|C) \qquad (2.8)$$

where C_m denotes choice sets containing m alternatives, $P(C_m|J_q)$ is the probability of choosing a choice set of size m from the set of all possible choice sets and $P(C|C_m)$ is the probability of selecting a particular choice set from the set of choice sets containing m alternatives. The model of equation (2.8) has been designated the distributed choice set (DCS) model.

A special case of the DCS model is when consideration is restricted to the M sets associated with C_1 and the single set C_{M_q} (= Ω_{M_q}); that is, each individual is assumed to either be captive to an alternative ($j = 1 \mid C_1$, $j = 2 \mid C_1$, .., $j = M \mid C_1$) or able to select from the full objectively determined choice set. The analyst, however, does not know to which group the individual belongs. Ben-Akiva and Swait (1984) have recently demonstrated that the captivity assumption, coupled with the set of assumptions concomitant with the derivation of a logit model, leads to a model of the form:

$$\text{Prob}\{I_q = j\} = \frac{1}{1 + \sum_{i \in \Omega_{M_q}} \exp(R_{qi}\iota)} \left(\exp(R_{qj}\iota) + \frac{\exp(\overline{V}_{qj})}{\sum_{i \in \Omega_{M_q}} \exp(\overline{V}_{qi})} \right) \tag{2.9}$$

where ι represents a vector of fixed unknown parameters, R_{qi} a row vector of variables describing the potential for individual q to be captive to alternative i, and Ω_{M_q} is the objectively determined choice set of individual q. By setting $\theta_i = \exp(R_{qi}\iota)$ and $\theta = \sum_i \theta_{qi}$, equation (2.9) can be written as:

$$\text{Prob}\{I_q = j\} = \frac{\theta_j}{1 + \theta} + \frac{1}{1 + \theta} \frac{\exp(\overline{V}_{qj})}{\sum_{i \in \Omega_{M_q}} \exp(\overline{V}_{qi})} \tag{2.10}$$

$$= \frac{\theta_j \sum_{i \in \Omega_{M_q}} \exp(\overline{V}_{qi}) + \exp(\overline{V}_{qj})}{1 + \theta \sum_{i \in \Omega_{M_q}} \exp(\overline{V}_{qi})} \qquad (2.11)$$

which is the dogit model of Gaudry and Dagenais (1979). Gaudry (see the previously cited work) and Ben-Akiva (1977) were the first to forge the connection between captivity and the dogit model. When θ_i is interpreted as indicating the likelihood that an individual is captive to the ith alternative, and a normalisation imposed so that $\theta_M = 1$, then it can easily be seen that equation (2.10) corresponds to Manski's general choice model of equation (2.7) (noting that $P(j \mid C_1^j) = 1$). The function exp $(R_{qi}\,\iota\,)$ in equation (2.9) merely represents an expansion of the captivity parameter. This has led to Swait and Ben-Akiva (1985) to dubb the model of equation (2.9) 'the parameterised logit captivity (PLC) model'.

Swait and Ben-Akiva (1985) also estimated the PLC model in a mode choice context using data from Sao Paulo, Brazil. They demonstrate theoretically and empirically that if a situation indeed exists in which choice sets consist of either one alternative or all alternatives, parameter estimates obtained from a simple MNL model will be downward biased. The high incidence of captivity in mode choice situations has been demonstrated by Ampt et. al (1985) among others. It is unlikely, however, that captivity will be nearly as pervasive in destination choice situations. This calls for a more general consideration of choice set formation than that provided by the PLC and dogit models.

Much research has appeared in the geography and psychology literature concerning spatial perceptual fields. Primarily this research has involved collecting, through application of a variety of techniques, information on an individual's 'awareness space'. A common finding is that individual's spatial knowledge extends toward the city centre in an elliptical or wedge shaped fashion. Another finding is that the area of spatial knowledge tends to be positively related to the individual's socio economic status. Also awareness space varies by residence location and is affected by landscape features such as rivers and railway lines (Lee 1954).

Recently some authors (Meyer 1980, Van der Heijden and Timmermans 1984) have attempted to marry this line of research with economic search theory to obtain operational models of spatial knowledge more in harmony with the precise requirements of a discrete choice modelling framework. Van der Heijden and Timmermans (1984) estimated a model conforming to the first probability term in equation (2.6)*. Data on familiarity with twenty-four shopping centres was collected and compressed into a series of binary variables. The probability that an individual was aware of the ith shopping centre was found to decline with distance to the centre and increase with shopping centre size. It was also negatively affected by the existence of an intervening shopping centre between the individual's residence and centre i.

There seems to be a least two major problems in pursuing this line of research. One concerns the relationship between the level of knowledge and the choice sets used in random utility models. What level of knowledge is necessary for an alternative to be regarded by the analyst as being in the choice set? Are all known alternatives, considered alternatives, and therefore properly deemed to habit the choice set? An important distinction in considering these issues is between knowledge gained through passive gathering of information and an active search process instituted by the individual.

---

* The modelling approach replicates that used in an early study of migration patterns by Brown et al. 1977.

The individual is constantly bombarded with information about alternative destinations, through radio, television, newspapers, billboards, personal contacts and the individual's own travels. Through these channels the individual is a passive recipient of information. The stock of knowledge collected in this way has been referred to as an individual's awareness space (Brown and Moore 1970).

Awareness space would seem to constitute the basis for active search behaviour. Active search can take the form of the individual actually visiting the destination or otherwise seeking information regarding salient destination attributes. When a destination has been actively searched it can safely be regarded as habiting the choice set, but is this also true for destinations for which even large amounts of passively gained knowledge are held?

A closely connected problem stems from treating knowledge, and hence choice set inclusion, as a binary concept. In reality knowledge takes on many different shades. It is multi-dimensional and continuous, not discrete. For destination choice two aspects of knowledge are of fundamental importance: knowledge of the location and knowledge of the destination attributes. Moreover, individuals possess knowledge in these areas to varying extents. For instance, one individual may possess very good knowledge about a destination attribute. Another, while still able to 'guestimate' a mean value, may be quite unsure as to the true value. This implies that knowledge constraints impact both the LHS and RHS of equation (2.1). To adequately introduce knowledge constraints into random utility analysis it is necessary that the choice set effect and the effect on parameterisation and variable values in the satisfaction (utility) function be treated in an integrated manner.

Before concluding this section it is useful to discuss briefly the objective assignment of destination alternatives to the choice set. It will be noted that even the objectively defined choice set, Ω_M, was subscripted by q in equations (2.6) and (2.11). This is in recognition of research into space/time prisms that originally received prominence in the geography literature and filtered through to transport mainly due to the efforts of members of Transport Studies Unit, Oxford University (Jones et al. 1983). Space/time prisms have been applied to shopping destination choice set identification by Landau et al. (1982). Place constraints were conditioned by the locations of shopping destinations, the home and, where applicable, school starting hours and shop opening times. These constraints, when considered in conjunction with the transport system, enabled delineation of the area of spatial reach for any person at any point in time, which effectively defined the individual's objective choice set. Prior to imposition of these restrictions 35 alternative shopping locations were (assumed) available to each individual. With the restrictions the average number declined to 28 locations, but some individuals had as few as 10 locations in the choice set. A MNL model developed with the restricted choice sets yielded slightly better predictions than for an unconstrained choice set model.

The research of Landau et al. and more generally Jones et al., points to a need to consider shopping within a daily stream of participation in activities. Not only will shopping destination choice be affected by the individual's current location and time commitments, but also participation in past and planned future activities. Furthermore, the individual's current location cannot automatically be assumed to be the place of residence, as has been explicitly or implicitly assumed in most shopping destination choice models. These models may be termed 'a situational'. Dynamic models that are situation specific are considered in the next section.

5. QUASI-DYNAMIC MODELS OF SHOPPING DESTINATION CHOICE BEHAVIOUR

A major impetus behind the study of shopping behaviour within the context of a daily activity schedule was results from surveys, designed to measure the effects of the US energy crisis of the mid-1970s, which showed that the most significant short-run effect was an increase in the level of multi-trip and multi-purpose travel (National Opinion Research Centre 1974, cited in Lerman 1979). This effect far outweighed traditionally studied responses such as mode and home-based destination switching, yet no model available at that time was able to capture adequately this pattern of change in travel behaviour. More generally, recent research has indicated that the proportion of multi-trip and multi-purpose tours in urban travel is significant and increasing (Hummon and Burns 1981, Kitamura 1983, Kitamura et al. 1981, Oster 1978).

The conventional travel demand modelling process effectively recognises two types of shopping trips: trips which originate or terminate at the home of the traveller (home-based trips) and trips with neither origin nor destination at the home (non-home-based trips). Models for each trip type are developed independently at each stage in the estimation process. That is, trips observed in the raw data are classified as being home-based or non-home-based and then used in model estimation. The weakness in this process is possibly most transparent in a mode choice context. From an intuitive standpoint, choice of mode (e.g. car driver) for the first trip on a tour is likely to significantly influence, even logically constrain, mode choice for subsequent trips on the tour, yet for independently developed home and non-home-based models no mechanisms exist for this to occur. From the summary of research by Landau et al. (1982) provided in the previous section, it can be seen that similar considerations apply to successive choices of destinations within a tour. Empirical evidence demonstrating the interdependence between trips on a tour can be found in Kitamura et al. 1981, Kitamura 1983 and Graham and Ogden 1979.

It is apparent that the conventional modelling process as applied to different trips on a tour lacks behavioural plausibility. In this section models are reviewed which operate from the perspective of entire travel patterns rather than individual trips. Firstly models which simultaneously consider entire travel pattern alternatives are reviewed. After this models which sequentially analyse travel and activity choices are examined.

5.1 SIMULTANEOUS MODELS OF TRAVEL PATTERNS

It is a simple matter to expand the utility formulation of destination choice, presented in Section 3 of this chapter to encapsulate entire travel patterns. Here the unit of analysis may be taken as an individual's travel pattern over some fixed time period. It may be postulated that the total utility (V_{tp}) obtained by an individual from pursuing a particular travel pattern, comprises a deterministic component ($\overline{V}_{tp}$) and a random component (ϵ_{tp}):

$$V_{(tp)} = \overline{V}_{(tp)} + \epsilon_{(tp)} \qquad (2.12)$$

Given that the $\epsilon_{(tp)}$ are distributed extreme value type 1 a MNL model may be derived, with the dependent variable compositely defined by combinations of modes, destinations and activities as included in the travel patterns.

The pioneering study along these lines was by Adler and Ben-Akiva (1979). These authors, using conventional transportation study data, assumed that $\overline{V}_{tp}$ was dependent upon the total number of destinations in the travel pattern (N_{tp}), the total number of destinations divided by the number of tours in the travel pattern (N_{tp}/T_{tp}), a vector of transport level of service variables (D_{tp}), a vector of variables measuring the attractiveness of destinations in the travel pattern (A_{tp}), and a vector of socio-economic variables (O):

$$\overline{V}_{tp} = \overline{V}_{tp}(N_{tp}, N_{tp}/T_{tp}, D_{tp}, A_{tp}, O) \qquad (2.13)$$

The terms N_{tp} and N_{tp}/T_{tp} were included to measure the 'scheduling convenience' associated with a travel pattern. In particular, Adler and Ben-Akiva hypothesised that in the absence of the constraining influence of travel, maximum utility would be attained by pursuing activities at spatially separate locations and at different times. The latter is due to the fact that activity needs will rarely temporally coincide. From this viewpoint maximum utility will result from pursuing each activity on a different tour, and in general utility will decrease as N_{tp}/T_{tp} increases. Similarly, given a set of activity needs and disregarding travel, maximum utility would be attained by pursuing each activity at the site most appropriate for that activity, rather than consolidating trips at the one location. Therefore as the number of visited destinations increases so will utility.

In the analysis conducted by Adler and Ben-Akiva both hypotheses concerning 'scheduling convenience' could not be rejected at usual levels of statistical significance. Travel times and costs were also found to be statistically significant as were most of the destination attractiveness variables and certain socio-economic descriptors.

The principal problems of the simultaneous approach to modelling travel patterns lie not in its theoretical base, which is quite general, but in application. Invariably, enumerating all possible travel patterns for each household leads to gargantuan choice sets with elements that are sometimes very closely related. Even though it is not necessary to enumerate completely all alternatives when estimating an MNL model some travel patterns will tend to be more closely related than others leading to potential violation of the IIA assumption used to derive the MNL model form. Furthermore, no means are provided for distinguishing between those travel patterns actually considered by households from the ones that are in principle available (Horowitz 1980). An alternative approach is to sequentially model travel patterns.

5.2 SEQUENTIAL MODELS OF TRAVEL PATTERNS

It is important to recognise initially that sequential modelling of travel patterns does not necessarily imply a sequential decision structure being followed by the individual. Just as it is now known (despite earlier confusion) that the nested modelling of mode choice within destination choice does not imply anything about the sequencing of these decisions by the individual, so Kitamura (1983) has shown the equivalence of sequential and simultaneous approaches to modelling travel patterns. Let $\overline{X}_{(n)} \equiv \{\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_n\}$ represent a sequence of events. Without loss of generality the probability of observing this event sequence can be expressed as:

$$\text{Prob}\ (\overline{X}_{(n)}) = \prod_{\ell=1}^{n} \text{Prob}\ (\overline{X}_\ell | \overline{X}_{(\ell-1)}) \qquad (2.14)$$

with $\overline{X}_{(0)} = \{\ \}$. The LHS of equation (2.14) is representative of a simultaneous modelling structure and the RHS, of a sequential modelling structure.

Nor is it correct to infer that the backward conditioning evident in the probability terms on the RHS of equation (2.14) precludes planning behaviour by the individual. Defining $\overline{X}_{(n*)}$ as $\{\overline{X}_r, \overline{X}_{r+1}, \ldots, \overline{X}_n\}$ and using Bayes' theorem, then:

$$\text{Prob}\ (\overline{X}_{r-1} | \overline{X}_{(n*)}) = \left[\prod_{\ell=1}^{n} \text{Prob}\ (\overline{X}_\ell | \overline{X}_{(\ell-1)})\right] \div \left[\sum_{\overline{X}_{\ell-1}} \prod_{\ell-1}^{n} \text{Prob}\ (\overline{X}_\ell | \overline{X}_{(\ell-1)})\right] \qquad (2.15)$$

The implication is that a choice conditioned on the past is also conditioned on the future (Kitamura 1983 p. 14).

The drawback in sequential modelling of travel patterns, then, is not because restrictive assumptions necessarily need to be imposed, but because of the difficulty in adequately representing the backward conditioning of activity participation. Effectively a sequential structure transfers the problem of modelling complex travel patterns from the dependent variable, that is from delineation of vast choice sets, to the set of independent variables. Even when only a narrow band of activity participation decisions are considered, the chain of conditionality, as shown in equation (2.14), soon becomes quite complicated. Complexity is substantially increased once a spatial component is introduced.

A useful framework for modelling trip chaining behaviour is provided by Markov and Markov renewal processes. A simple Markov model assumes a system comprising a finite number of states together with a probability law of moving from state to state. In a travel pattern context the 'states' may refer to land use types, geographical zones, etc. A matrix, P, is constructed showing the probabilities of moving from one state (i) to another (j). The probabilities, termed transitional probabilities, may be generated from the relative frequencies of travel between origin-destination (O-D) pairs at a point in time.

Stated formally,

$$\text{Prob}(\overline{X}_i = S_k^* | X_{i-1} = S_j^*) = \overline{p}_{jk} \qquad (2.16)$$

where, $\overline{X}_i$ denotes the state of the system at time point i, S_1^*, S_2^*, ..., S_w^* denote the system states, $\overline{p}_{jk}$ is the transition probability from S_j^* to S_k^*. Note that for the first order Markov chain of equation (2.16) $\overline{p}_{jk}$ is assumed free of i. That is, the process is memoryless with the probability of moving from j to k being independent of the states the traveller visited before j.

Let $\overline{X}_0$ denote the starting state of the system with probabilities given by

$$\overline{p}_k^{(0)} = \text{Prob}(\overline{X}_0 = S_k^*) \qquad (2.17)$$

Also $\overline{p}_k^{(n)}$ is the probability of being in state S_k^* after n steps. Then it can be shown that

$$P^{(n)} = P^{(n-1)} P \qquad (2.18)$$

where $P^{(n)} = (\overline{p}_1^{(n)}, \overline{p}_2^{(n)}, \ldots, \overline{p}_w^{(n)})$ and P is the matrix of transition probabilities, which are constant over time.

If P is regular (i.e. some power of P has all positive entries) then it is possible to eventually get from state S_k^* for any pair (j, k), and repeated matrix multiplication will yield a stationary (or equilibrium) distribution which gives the probabilities of being in the respective states after many transitions have evolved. Such a matrix is termed the A matrix, when

$$A = P^n \qquad (2.19)$$

A particular a_j can then be interpreted as the expected percentage of travellers which will be found in state j at a randomly selected point in time.

If a state is designated impossible to leave once entered (i.e. $\overline{p}_{ii} = 1$ and $\overline{p}_{ij} = 0$ for $j \neq i$) then the state is termed an absorbing state and an absorbing Markov chain can be formed. In the case where it is possible to eventually get to an absorbing state from every other state, the Markov chain will end up in an absorbing state. By operating upon this modified P matrix, with 'home' designated as the absorbing state (i.e. once the home state is re-entered it is impossible to leave), another matrix can be formed, the elements of which provide the expected number of trips made on each tour.

Simple Markov models of the form outlined have been used in descriptive studies by Marble (1964), Horton and Shuldiner (1967), Horton and Wagner (1968), Hemmens (1970) and Wheeler (1972). A major advance by Lerman (1979) was to relate the transitional probabilities to exogenous factors. This general approach has since been followed by Ben-Akiva et al. (1979) and Kitamura and Kemanshah (1983 and 1984).

Lerman utilised a semi-Markov process which represents an amplification of the basic Markov model by permitting incorporation of a time dimension. Formally a semi-Markov process is defined by:

$$P(t) = P\,F(t) \qquad (2.20)$$

where, P(t) is the matrix of probabilities of a state j to state k transition during time period (0, t), P is the matrix of transition probabilities of going from state j to state k and F(t) is a matrix with elements $f_{jk}(t)$ which represent the time of departing from j to k, given that the system is in state j and the next state is k, in the time period (0, t).

Four general transition states were recognised by Lerman; home (which is treated as a special state in that it is defined as the traveller's initial location and once left is never re-entered), destination d^H which is reached from home by mode m^H, destination d^{NH} which is reached from some non-home destination by mode m^{NH}, and another home state for subsequent arrivals and departures from this location. Corresponding to these states are three departure time distributions; $f^H_{m^H d^H}(t)$ - the distribution of the time of the first departure from home to mode/destination combination $m^H d^H$, $f^{\overline{H}}_{m^H d^H}(t)$ - the time distribution of subsequent home departures to mode/destination combination $m^H d^H$, and $f^{NH}_{m^{NH}d^{NH}}(t)$ – the distribution of departure times from non-home locations to mode/destination combination $m^{NH}d^{NH}$. Two logit models were developed to determine state transitional probabilities, one for home to destination travel, the other for travel originating from non-home locations. Gamma and modified exponential distributions were used to approximate the departure time functions. These functions define the frequency of travel. They are developed by Lerman outside the utility maximising framework and are assumed to be unaffected by accessibility and other factors associated with m or d.

A comparison of equations (2.20) and (2.16) with (2.14) reveals that the chain of conditionality is summarily truncated in first order Markov processes. Accordingly, many authors have expressed doubts about the applicability of memoryless Markov models to data on complex travel movements (e.g. Jones 1976, Hanson 1979, Hemmens 1966). Recent work by Kitamura (1983) and Kitamura and Kemanshah (1983 and 1984) has specifically addressed this aspect of the Markov model. Their work also represents an advance on Lerman's research in that there is differentiation of non-home activity types.

In a preliminary exploratory study to model development, Kitamura (1983) found that the stationarity and history-independent assumptions underlying the simple Markov chain were untenable. In particular, Kitamura found that the strength of linkages between activities was dependent both on their position within the trip chain and the history of previous activities pursued in the chain. There was a strong tendency for activities pursued within a chain to be relatively homogeneous. Nevertheless, Kitamura was able to obtain a good representation of observed activity linkages by formulating the conditional probabilities as:

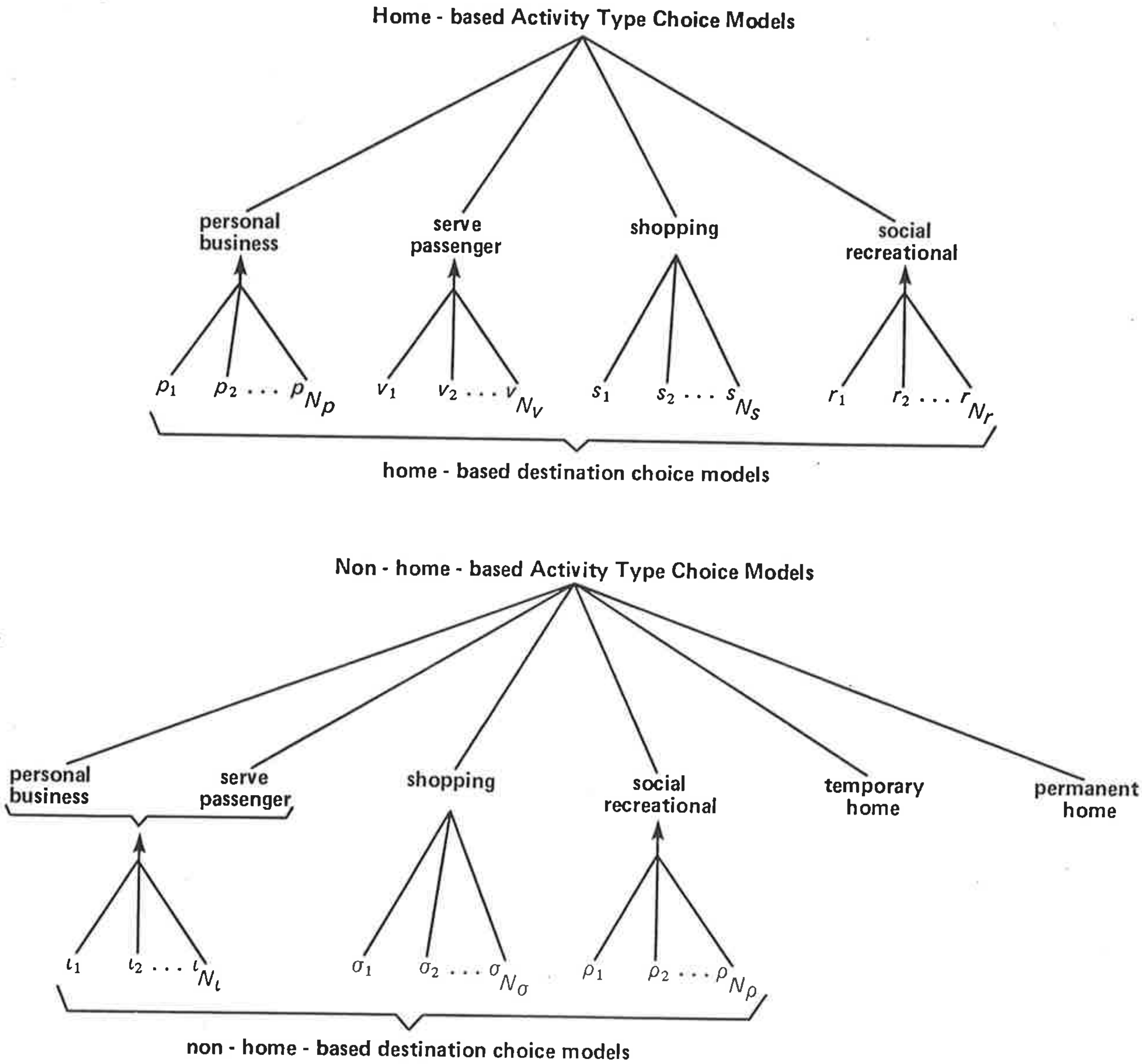
$$Pr(\overline{X}_{n+1} | \overline{X}_1, \overline{X}_2, \ldots, \overline{X}_n = Pr(\overline{X}_{n+1} | \overline{X}_n; \overline{B}_{1n}, \overline{B}_{2n}, \ldots, \overline{B}_{Kn}) \quad (2.21)$$

where $\overline{B}_{jn}$ is a binary (0 - 1) variable, which indicates whether activity type j has been pursued in the chain by the nth transition, and K is the number of activity categories used to represent the history of the chain.

Kitamura and Kemanshah - K & K - (1983 and 1984) were able to use this relatively simple depiction of activity participation history in a subsequent modelling effort. A schematic description of their modelling system for the transitional probabilities is shown in Figure 2.2. Four activity types that could be selected from home were recognised; personal business, serve passenger, shopping and social/recreational. An additional two activity types, temporary home and permanent home could be reached from non-home activity types, the latter acting as an absorbing Markov state. Nested within the activity type models are a number of destination choice models. Inclusive values were calculated from these models and passed on to the activity type choice models; but the associated coefficients mostly took on values statistically insignificant from zero and/or lying outside the 0 - 1 interval. Selection of shopping as an activity, in particular, appeared to be unaffected by accessibility to shopping opportunities.

A unique feature of the models was that the transitional probabilities were made time-dependent. This was achieved by including time of day as an independent variable and in combination with other variables, such as travel times and destination attractiveness descriptors, in the utility expressions of the logit models. The models provided strong evidence of temporal variations in destination choice and activity participation behaviour. Additionally, the utility expressions for non-home destination choice contained a variable measuring home-destination travel time, thus building into the models an element of individual forward planning *. This variable was

* A more general attempt to embed concepts of individual planning into models of destination choice with trip chaining can be found in Kitamura (1984).



Notes

- $p_1, p_2, \ldots, p_{N_p}$ refer to personal business destinations accessible from home
- $v_1, v_2, \ldots, v_{N_v}$ refer to serve passenger destinations accessible from home
- $s_1, s_2, \ldots, s_{N_s}$ refer to shopping destinations accessible from home
- $r_1, r_2, \ldots, r_{N_r}$ refer to social/recreational destinations accessible from home
- Greek letter equivalents of the above refer to destinations accessible from non - home locations
- arrows indicate the conveyance of inclusive value terms into the activity choice models

Fig. 2.2 — Schema of Modelling Structure adopted by Kitamura and Kermanshah 1984

found to be at least as significant as the traditional origin-destination travel time. The activity history variables of the type previously discussed were important in influencing activity choice, but not destination choice.

Kitamura and Lam (1983) discuss linking the transitional probability matrix outlined above to a sojourn time distribution matrix which is current activity, but not past activity, future activity, time of day or location, dependent. Despite the simplicity of this structure, the model demonstrates many empirically plausible and useful properties.

Although the work of Kitamura and Kemanshah represents the most advanced study of its type, the examination of historical dependencies in destination choice behaviour was rather limited and may be criticised in two major areas:

(i) Only a narrow range of possible historical dependencies were examined. In the context of destination choice Kitamura and Kemanshah only examined the possibility that a future destination choice was dependent on the current destination and activity choice and on past activity choices. There remains the strong possibility, for instance, that a future destination choice for an activity is related to past destination choices for that activity (for evidence, see Miller and O'Kelly 1982). Further, the broad trip purpose categories used by Kitamura and Kemanshah and other similar studies may have served to mask detailed historical dependencies in behaviour. For example, observed destination choice behaviour may appear to be historically independent if all shopping is considered as one category, with dependencies only revealed when grocery shopping is considered in isolation. Finally, the data used by Kitamura and Kemanshah restricted the examination of historical dependency to a maximum period of one day. It may be that cycles in destination choice behaviour stretch over a longer period.

(ii) Some of the statistical techniques used by Kitamura and Kemanshah are suspect. Two issues requiring attention are the use of endogenous variables in one stage of the modelling process as exogenous variables at another stage (a difficulty recognised by the authors) and the failure to adequately distinguish between heterogeneity, non-stationarity and state dependence effects in modelling discrete choices over time.

These criticisms are not pursued here. However, in Chapter 8 brief coverage is given to the wider econometric literature on the modelling of discrete choices over time. This literature serves to expose the limitations of approaches favoured in the transport research community.

6. INTEGRATED MODELS OF SHOPPING DESTINATION CHOICE AND SHOPPING EXPENDITURE

Another area of recent development in the modelling of shopping behaviour has been linking models of shopping destination choice to models of shopping expenditure. Interest in this area would seem to have stemmed from two major sources. One concerns the critique by some transport researchers (especially those from the TSU, Oxford University) that discrete travel choice models ignore important aspects of activities conducted at trip ends (e.g. Jones et al. 1983, p. 6; see also Damm 1984, p. 249). The other is more practical in orientation. With increasing levels of retailing competition there has developed a need for planners and retailers alike to obtain more accurate predictions of retailing activity under alternative scenarios. For instance, among planners, there has been recent interest in the effects of transport associated policy measures on retail sales, particularly in the CBD (Atherton and Eder 1982, Kern and Lerman 1982, Lundon and Coogan 1985), presumably to ensure continued financial viability for existing centres. Retailers also have shown keen interest in obtaining accurate predictions of the effect of site location and centre characteristics on patronage and turnover.

Two early studies that considered shopping travel behaviour in conjunction with shopping expenditure were Rushton et al. (1967) and Huff (1963). As pointed out in Chapter 1, developments of the Huff model still enjoy wide use. From equation (1.5), the model basically involves weighting the destination choice probabilities by consumers' anticipated expenditure levels where the latter are made a function of socio-economic characteristics. The output is expected expenditure by each consumer at each centre. The values can be summed over consumers to obtain predicted retail sales at each centre. Later variants of the Huff model include the expenditure weighted multiplicative competitive interaction (MC1) model (e.g. Nakanishi and Cooper 1974) and the multinomial logit model (e.g. Malhotra 1984).

Vickerman and Barmby (1984) is the first published work recognising two important aspects of model systems accounting for both patronage and expenditure decisions. * Firstly, given a utility based theoretical foundation for the system, the model used for the patronage decision and that used for the expenditure decision are not unrelated, as implicit in the Huff formulation. Secondly, the system statistically falls into a general class of econometric models dealing with interconnected discrete and continuous choices.

---

* The work detailed in Chapter 3 of this thesis also recognises these aspects of model systems accounting for both patronage and expenditure decisions and was developed independently from that of Vickerman and Barmby.

The main thrust of Vickerman's and Barmby's research was in the **number** of shopping trips made by the household and expenditure. To do this they set up a theoretical model only loosely based on the tenets of economic utility theory. Their formal exposition consists of four equations. They begin with a trip generation relationship of the form:

$$T_q = f_1(p_{iq}, Y_q, \bar{s}_{iq\ell}) \tag{2.22}$$

where T_q is, for a given period, the number of trips made by individual q, p_{iq} is the price of good i for individual q, Y is income and $\bar{s}_{iq}$ is the shopping costs associated with purchasing good i. In turn, $\bar{s}_{iq}$ can be specified as:

$$\bar{s}_{iq\ell} = \bar{s}_{iq\ell}(a_{iq}, b_{iq\ell})$$

where a_{iq} is the acquisition cost of purchasing good i and $b_{iq\ell}$ is the travel cost for individual q living at location ℓ. Acquisition costs are meant to cover individuals' time spent in search and other incidental costs.

The budget constraint faced by the individual is:

$$B_{q\ell} = \sum_{i}^{N} (p_{iq} + \bar{s}_{iq\ell}(a_{iq}, b_{iq\ell}))\, C_{iq} \tag{2.23}$$

where C_{iq} is the quantity of good i purchased by individual q. From equation (2.23) expenditure on shopping goods by individual q is:

$$E_q = \sum_{i}^{N} p_{iq}\, C_{iq}$$

or,

$$E_q = f_2(B_{q\ell}, p_{1q}, \ldots, p_{Nq}, t_{1q\ell}, \ldots, t_{Nq\ell}, a_{1q}, \ldots, a_{Nq}) \tag{2.24}$$

Barmby and Vickerman hypothesize that the level of trip-making is the result of a 'constrained planning process whereby the individual undertakes the minimum number of trips that will still enable him to achieve the desired level of expenditure and be consistent with the budget constraint'. They specify the resulting level of trip-making as:

$$T_q = f_3(E_q, t_{1q\ell}, \ldots, t_{Nq\ell}) \quad (2.25)$$

Equations (2.24) and (2.25) represent the link between expenditure and trip-making. Empirically this link can be handled by two stage generalised least squares.

Even Vickerman and Barmby recognise some problems inherent in the formulation of equations (2.22) - (2.25). Important deficiencies are that the theoretical analysis is partial in orientation (there is no formal interaction between shopping expenditures and the prices attached to other activities) and that travel costs are treated in a proportionate manner to quantities of goods purchased. (Clearly the spatial distribution of opportunities need bear no relationship to quantities of goods purchased.) Their empirical work did not differentiate between trips of different destination and indeed it would be difficult to incorporate this generalization within the approach adopted. Furthermore the simultaneous regression techniques used exhibit well-known weaknesses when applied to discrete choices. The authors justify the use of regression by arguing, '... the complexity of the way in which trips can be ordered and the rather different ways in which their significance could be assessed in relation to a utility-maximising hypothesis would also pose problems for a simple model of discrete choice' (p. 120). I will show, particularly in Chapters 3 and 7, that a discrete/continuous model system based on the tenets of utility maximising theory can readily be constructed to account for shopping behaviour.

7. CONCLUDING COMMENTS

Later chapters in this thesis contribute advances in most of the areas covered in the foregoing review. However, from the review two areas emerge as being particularly weak. Firstly, immaturity is evident in the theoretical underpinnings provided for the models. Secondly, comparatively little work has been conducted in linking shopping destination choice and expenditure decisions. It so happens that a sophisticated theoretical specification of how destination choices are made also establishes a link between these decisions and those to do with shopping expenditure. This theory is set out in Chapter 3. Estimation of the theoretically established link between these decisions involves appropriate statistical recognition of the relationship. Statistical estimation of an integrated store choice/shopping expenditure model is contained in Chapter 7.

Two further areas covered in the review receive detailed attention in forthcoming chapters. In Chapter 5 variations in reported choice sets are examined, together with the impact which choice set specification exerts when estimating discrete shopping destination choice models. In Chapter 6 a model of food shopping travel patterns is described. The approach is similar to that used by Adler and Ben-Akiva, but the study is conducted at a micro level, specifically confined to urban food shopping. The micro level focus of this study, as with other aspects of this thesis, aligns with Hanson's (1979) suggestion:

> 'It is my opinion that attempting to build a single, all encompassing super model of travel linkages would be inappropriate at this time. A more reasonable approach is to construct any number of more modestly conceived models aimed at addressing selected aspects of urban travel linkages.' (p. 95)

CHAPTER 3

A THEORETICAL MODEL OF SHOPPING DESTINATION CHOICE AND ITS RELATIONSHIP WITH SHOPPING EXPENDITURE

1. INTRODUCTION

Research presented in this chapter has the general objective of strengthening the theoretical underpinnings of discrete shopping destination choice models. Two specific aims are also evident.

The first specific aim stems from recent criticism in the transport literature of discrete choice travel models on the grounds that they ignore aspects of activities conducted at trip ends. It is argued by these critics that these activities, which are the cause for travel, are much neglected in current travel modelling schemes. For example Damm (1984) states: 'Even in the context of recent 'behavioural' modelling of transportation-related choices, only a small amount of time has been spent trying to explore the highly interrelated nature of people's activities and the resulting consequences for travel behaviour'. Jones et al. (1983) in this context provide a more specific critique of discrete choice travel models (p. 6): 'The models do not provide a means by which the interrelationships between travel and non-travel aspects of life may be better understood. Technological advances relating to certain activities ... may affect activity participation and so indirectly travel patterns; conversely, transport policies may have important non-travel impacts'.

It should be recognised that current approaches to modelling travel choices do not altogether ignore trip end activity participation. The traditional stratification of these models is in explicit recognition of the derived nature of travel demand. Economists have long hypothesised, from general theory, certain aspects of the relationship between the demand for a final activity and the derived demand for an intermediate activity. In particular the elasticity of derived demand is hypothesized to be an increasing function of the elasticity of demand for the final activity, the ratio of intermediate to final activity cost and the availability of supply of final activity (see Hicks 1936 for formal proofs). Stratification of travel choice models by trip purpose permits the parameters to be affected by the **type** of activity pursued at the trip destination. No link has been forged, however, between travel and the **intensity** of activity participation. The first specific aim is to demonstrate from economic theory the links which exist between one particular travel decision, shopping destination choice, and the degree of participation in this activity measured by shopping expenditure.

The second specific aim is related to the Huff model. In particular, many users of this model genre have advocated a utility interpretation for the F_{iq} functions. Indeed, this suggestion is evident in Huff's original work. It is an intention of this chapter to demonstrate that the treatment of expenditure in Huff-type shopping models is inconsistent with a utility based interpretation for the F_{iq} functions. It will also be shown how empirical models consistent with utility maximisation theory can be derived. These models represent a generalisation of the generalised Huff model.

Taking a wider view, this chapter forges the link between two hitherto disparate approaches to examining shopping behaviour. One approach characterised by discrete choice shopping models has involved analysing the decision of where to shop in isolation of how much to spend (see, for example, Domencich and McFadden 1975, Recker and Kostyniuk 1978, Koppelman and Hauser 1977, McCarthy 1979, Gautschi

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 (see, for example, Curhan 1972, Guy 1984 and Morey 1980). To the extent these choices are interrelated these models will be less than complete and results from them may be biased. From an information viewpoint it is obviously beneficial for developers and planners to know both the number of persons using a store and expenditure at that store.

The remainder of this chapter takes on the shape of a funnel. Most of the objectives enunciated above are met in the following section which proposes a general theoretical framework for analysing shopping destination and expenditure choices. More specificity than this, however, is required before obtaining an estimable model. The specifics are dealt with in Sections 3-5.

2. SHOPPING CENTRE AND EXPENDITURE CHOICE: THE GENERAL THEORETICAL FRAMEWORK

At the core of theory presented in this section is the assumption that an average consumer 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 and the analysis conducted at a household level, provided the household has a single dictatorial decision maker. A general form of the consumer's utility function may be written as:

$$U = U\ (G,\ B_1,\ B_2,\ \ldots,\ B_N,\ \overline{Z},\ L) \qquad (3.1)$$

where G is a vector $(g_1, g_2, \ldots, g_N)$ representing consumption of shopping items from centres 1, 2, ..., N, respectively, B_i is a vector $(b_{i1}, b_{i2}, \ldots, b_{iK})$ of quality variables associated with the consumption of shopping items from the ith centre, $\overline{Z}$ is the Hicksian composite commodity encapsulating consumption of other goods and L is leisure time.

Maximisation of utility is constrained by an income constraint and a time constraint:

$$Y = \sum_{i}^{N} p_i g_i + \overline{Z} + \sum_{i}^{N} \xi_i c_i \qquad (3.2)$$

$$L = T - \sum_{i}^{N} \xi_i t_i \qquad (3.3)$$

where p_i is an index of shopping prices at the ith destination alternative, $\xi_i = \xi_i\ (g_i)$ is an indicator function with $\xi_i = 1$ if $g_i > 0$ and $\xi_i = 0$ if $g_i = 0$, c_i is the cost of travel to the ith destination alternative, t_i is the time taken to travel to the ith destination alternative, Y is income and T is total time available. Note from the above, alternative i is strictly taken to be a centre/mode choice combination (since travel times and costs vary by alternative modes as well as centres). However, it is often conceptually easier to think of i solely in terms of centre choice. This is the practice adopted in presenting the model.

Truong and Hensher (1985) have identified the treatment of time in the equation system (3.1), (3.2) and (3.3) as following the Becker framework. It represents a development of the simple work leisure model (Robbins 1930, Train and McFadden 1978) in that independence is assumed between the income and time constraints. One reason for such independence may be a fixed working week. The result is that time cannot necessarily be traded for income at the wage rate.

Thus far the theory presented does not diverge greatly from classical consumer theory. Two primary modifications to this approach are now made. Firstly, it is explicitly recognised that although to the consumer U(.) is known with certainty, the analyst is only able to observe a portion of individual utility obtained from the consumption of commodities. Equation (3.1) can therefore, be rewritten from an analyst's perspective as:

$$U = U\ (G,\ B_1,\ B_2,\ \ldots,\ B_N,\ \overline{Z},\ L,\ \epsilon\ ) \tag{3.4}$$

where ϵ is a vector $(\epsilon_1, \epsilon_2, \ldots, \epsilon_N)$ of unobserved influences on utility which can be treated as random variables. This modification is in the spirit of random utility theory first formulated by Marschak (1959) and subsequently developed by McFadden (esp. 1975, 1978). Secondly, an element of discreteness is introduced in the model by assuming that in any time period the consumer only selects one destination for shopping purposes. This restricts the solution of the constrained maximisation problem to be such that $\overline{Z}$, L and **one** of the g_js 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, $\overline{Z}$ and L can be consumed in any quantities.

In general, two main reasons can be identified for the maximisation solution to be of the form alluded to above (Small and Rosen 1981). One is that the choice of some goods may be restricted to a small number of mutually exclusive varieties. Mutual exclusiveness may arise due to supply, institutional or logic constraints. A particularly common case is that goods may only be available in discrete quantities which are so large that most consumers can consume but one unit at any point or period of time. An example is housing tenure: within current housing markets a consumer cannot, for instance, rent a bathroom and own a kitchen; rather the rent/ownership decision can only be made with respect to an entire house. Other examples include college degrees and transport mode choice. For these goods the form of solution is effectively the result of an extra constraint operating on the maximisation process. This constraint may be written as:

$$g_i g_j = 0 \text{ for all } i \neq j \tag{3.5}$$

A second, possibly more pervasive, reason for discreteness, however, relates to the shape of the utility function. In this case, the consumer's preferences are such that only one of the g_is is selected at any time. This is assumed to be the cause of discreteness in the shopping destination/expenditure choice problem analysed in this thesis. In particular it is assumed that the consumer views alternative shopping destinations as perfect substitutes. Formally this implies non-concavity of the utility function leading to a choice between alternative corner solutions.

In obtaining optimal values of the g_is, $\overline{Z}$ and L the consumer can be thought of as applying a two stage maximisation process. The first stage involves redefinition of the utility function initially with $g_1 > 0$ and $g_2 = g_3 = \ldots = g_N = 0$, with concomitant simplification of the income and time budget constraints. If it is also assumed that if $g_i = 0$ then $\frac{\partial U}{\partial b_{i1}} = \frac{\partial U}{\partial b_{i2}} = \ldots = \frac{\partial U}{\partial b_{iK}} = 0$, the maximisation problem can be redefined as:

$$\max U_1 = U_1 \ (g_1, B_1, \overline{Z}, L, \epsilon) \tag{3.6a}$$

$$\text{subject to: } Y = p_1 g_1 + \overline{Z} + c_1 \tag{3.6b}$$

$$L = T - t_1 \tag{3.6c}$$

where U_1, is assumed to possess the usual properties of monotonicity, quasi-concavity and differentiability. The solution to (3.6a) - (3.6c) is a set of demand equations:

$$g_1^* = g_1^* \ (p_1, B_1, t_1, c_1, Y, \epsilon) \tag{3.7a}$$

$$\overline{Z_1^*} = \overline{Z_1^*}\ (p_1,\ B_1,\ t_1,\ c_1,\ Y,\ \epsilon\ ) \qquad (3.7b)$$

$$L_1^* = L_1^*\ (p_1,\ B_1,\ t_1,\ c_1,\ Y,\ \epsilon\ ) \qquad (3.7c)$$

This process is repeated for $g_2 > 0$, $g_1 = g_3 = \ldots = g_N = 0$ and so on. The U_i have been labelled **conditional** direct utility functions as they are conditional on choice i being made. Analogously g_i^*, $\overline{Z_i^*}$ and L_i^* are conditional demand functions. The second stage in the maximisation process then involves the consumer comparing utility levels from conditional maximisation and applying a rule: choose j iff $U_j\ (g_j^*,\ \overline{Z_j^*},\ L_j^*,\ B_j,\ \epsilon\ ) \quad U_i\ (g_i^*,\ Z_i^*,\ L_i^*,\ B_i,\ \epsilon\ )$ for all $i \neq j$.

For empirical applications, rather than working directly with the $U_i\ (g_i^*,\ \overline{Z_i^*},\ L_i^*,\ B_i,\ \epsilon\ )$, it is often convenient to substitute in the price vector, travel time and cost and income to obtain:

$$V_i = U_i\ (g_i^*,\ \overline{Z_i^*},\ L_i^*,\ B_i,\ \epsilon\ ) = V_i\ (\ p_i,\ B_i,\ t_i,\ c_i,\ Y,\ \epsilon\ ) \qquad (3.8)$$

where V_i is termed the conditional indirect utility function. Shopping destination j will be chosen if:

$$V_j\ (p_j,\ B_j,\ t_j,\ c_j,\ Y,\ \epsilon\ ) >$$

$$V_i\ (p_i,\ B_i,\ t_i,\ c_i,\ Y,\ \epsilon\ ) \text{ for all } i \neq j \qquad (3.9)$$

The V_i are the V_i encountered in equation (2.2) with the non-stochastic elements represented by $\overline{V_i}$ (equation (2.1)) and the F_i functions of the generalised Huff model (equation (1.5)). Note for certain functional forms, in comparing shopping destinations i and j, Y can be deleted. Also note these functions contain variables describing prices at destination i, other attractiveness variables associated with destination i, and travel times and costs to destination i - in other words, all the variables normally included in a behaviourally based destination choice model.

The convenience of working with the indirect utility function derives from the fact that demand equations which are consistent with consumer maximising behaviour can be obtained through simple differentiation of the V_i rather than explicitly solving the previously set out maximisation problem. In particular Roy (1942) has shown for an indirect utility function of the form $V = V(p_1, p_2, \ldots, p_N, Y)$ corresponding demand equations, C_i, can be derived by applying the identity $C_i = \dfrac{-\partial V/\partial p_i}{\partial V/\partial Y}$. As is proved in Appendix 3A this holds even when V is expanded to include travel time and cost, quality and error variables. This permits derivation of conditional demand equations for the consumption of shopping goods from the conditional indirect utility functions as follows:

$$g_i = -\frac{\partial V/\partial p_i}{\partial V/\partial Y} = g_i\,(p_i, B_i, t_i, c_i, Y, \epsilon) \qquad (3.10)$$

For i = 1 equation (3.10) is exactly the same as equation (3.7a). Equation (3.10) can be expressed in expenditure form by multiplying both sides by p_i:

$$p_i g_i^* = E = p_i g_i^*\,(p_i, B_i, t_i, c_i, Y, \epsilon) \qquad (3.11)$$

Equation (3.11) and the corresponding indirect utility function (equation (3.8)) establishes the link between shopping destination and expenditure choices.

To this point the analysis has been for one consumer. When a number of consumers are considered the conditional indirect utility functions and demand equations must also be subscripted by q; thus, V_i becomes V_{iq}, the conditional indirect utility function associated with shopping mode/destination alternative i and individual q, and E_i becomes E_{iq}.

Three aspects of the theoretical analysis this far warrant highlighting. Firstly, in modelling shopping destination choices, provided the V_{iq} are correctly specified, **the level of activity involvement directly shapes the consumer's travel decision**. This is evident from equation (3.8). This equation shows that the optimal level of activity involvement is embedded in the indirect utility functions used to model shopping (destination) travel choices. Secondly, **the travel environment** facing the consumer and his travel decisions **directly affect** his chosen (optimal) **level of activity involvement**. This is evident from equation (3.10). Notice the symmetry here between travel choices and activity participation choices. Finally, a major weakness has been exposed in Huff type models; namely, expenditure must, in general, be allowed to vary by attributes pertaining to shopping destinations, as well as across individuals. Any other formulation is inconsistent with a utility interpretation for this genus of models.

3. ISSUES IN THE SPECIFICATION OF THE QUALITY INDICES AND FORM OF THE CONDITIONAL INDIRECT UTILITY FUNCTIONS

In developing the model further to the stage that it is amenable to empirical estimation, it is necessary to select a particular method for inclusion of the quality variables, choose a functional form for the conditional indirect utility functions, decide on an appropriate structure for the error terms, and develop the mechanics of statistical estimation. In this Section the first two topic areas are discussed. These may be regarded as preliminary topics to the development of two procedures in Sections 4 and 5, which lead to estimable forms for the model system derived in Section 2. One procedure is based on the work of Hanemann (1984), the other utilises a full conditional indirect utility function approach. Statistical estimation issues for the integrated destination/expenditure choice system are primarily dealt with in Chapter 7.

3.1 QUALITY VARIABLE INCLUSION

In considering the quality variables it is initially convenient to compress the (b_{11}, b_{12}, ..., b_{1K}, b_{21}, ..., b_{2K}, ..., b_{iK}, ..., b_{NK}) into a single index for each i (i = 1, 2, ..., N). The index can be denoted $\psi_i = \psi_i(\gamma, B_i)$ with γ representing a vector of unknown fixed parameters ($\gamma = \gamma_1, \gamma_2, \ldots, \gamma_K$). As will be seen, for the purposes of research reported here, it is computationally convenient to assume that the ψ_is are exponentially multiplicative in the b_{ik}s (k = 1, 2, ..., K), such that:*

$$\psi_i = \exp\left(\sum_k \gamma_k b_{ik}\right) \tag{3.12}$$

At least two methods then exist for inclusion of the quality variables in the conditional indirect utility functions.

One method is to expand the constant terms in the conditional indirect utility expressions. As an example, for a simple linear specification of the conditional indirect utility functions:

$$V_i = \alpha_{1i} + \alpha_2 p_i + \alpha_3 Y + \epsilon_i \tag{3.13}$$

quality variables might be included by defining α_{1i} as:

$$\alpha_{1i} = \alpha_1^* + \log \exp\left(\sum_k \gamma_k b_{ik}\right) \tag{3.14}$$

Another, theoretically more elegant, method for inclusion of the quality variables is to follow the simple repackaging hypothesis of Fisher and Shell (1971). When

---

* The quality vectors b_i, i = 1, 2, ..., N, may be of different length and composition. Here they are implicitly standardised by constructing a global set of quality variables and including 0 values for alternatives where a particular quality variable is undefined.

there is perfect substitutability between the g_is a convenient general form of the direct utility function incorporating the simple repackaging hypothesis can be written as (Deaton and Muellbauer 1980a):

$$U = U^{*} \left(\Sigma\, g_i \psi_i,\ \overline{Z},\ L,\ \epsilon \right) \tag{3.15}$$

where U^* is a conventional bivariate utility function. It is easy to see diagrammatically in the two good case that a utility function such as (3.15) ensures a corner solution with at most one $g_i > 0$. * Denoting $h = \Sigma\, g_i \psi_i$ and totally differentiating (3.15) yields:

$$dU = \sum_i^2 \frac{\partial U}{\partial h} \psi_i dg_i + \sum_i^2 \frac{\partial U}{\partial h} g_i d\psi_i + \frac{\partial U}{\partial \overline{Z}} d\overline{Z} + \frac{\partial U}{\partial L} dL + \sum_i^2 \frac{\partial U}{\partial \epsilon_i} d\epsilon_i \tag{3.16}$$

By holding $\overline{Z}$ and L, the quality indices and error terms constant, equation (3.16) can be simplified to:

$$dU = \psi_1 \frac{\partial U}{\partial h} dg_1 + \psi_2 \frac{\partial U}{\partial h} dg_2 \tag{3.17}$$

with $dU = 0$

$$\frac{dg_1}{dg_2} = -\frac{\psi_2}{\psi_1} \tag{3.18}$$

which is the slope of the set of indifference curves. Since this slope is independent of g_1 and g_2 it is a straight line as depicted in Figure 3.1. The budget constraint for fixed expenditure, Y_g, on goods g_1 and g_2 is also shown. It is apparent that, except in the limiting case where $\frac{\psi_1}{\psi_2} = \frac{p_1}{p_2}$, causing the budget line and the indifference curves to take

---

* It is implicitly assumed throughout that grocery shopping is an essential activity implying that g_i is essential with respect to U_i and ensuring that the solution to unconditional maximisation will be such that one $g_i > 0$.

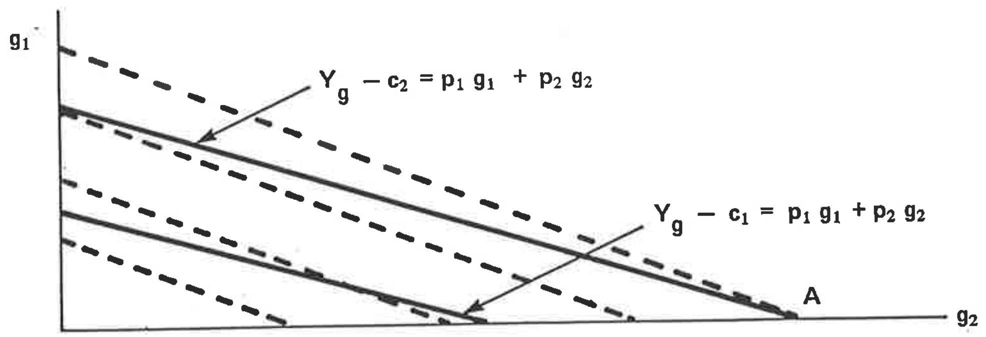


Fig 3.1 — Indifference curves (shown by dotted lines) and budget lines (shown by unbroken lines) for the unconditional utility function of equation (3.15) with N = 2.

Notes:
* Slope of indifference curves $= -\dfrac{\psi_1}{\psi_2}$
* c, and c_2 are, respectively, the travel costs associated with the consumption of g_1 and g_2
* Other notation is defined in surrounding text

on the same slope, the budget line will intersect the highest attainable indifference curve at a point where either $g_1 = 0$ or $g_2 = 0$. In Figure 3.1 this is point A.

The **conditional** direct utility function corresponding to the function of equation (3.15) can be expressed as:

$$U_i = U^* (g_i \psi_i, Z, L, \epsilon) \qquad (3.19)$$

and Muellbauer (1975) has shown the conditional indirect utility function is given by:

$$V_i = V^* (p_i / \psi_i, t_i, c_i, Y, \epsilon) \qquad (3.20)$$

Equation (3.20) together with the demand function derived from it, namely,

$$g_i = g_i (p_i / \psi_i, t_i, c_i, Y, \epsilon) \qquad (3.21)$$

form the basis of the model developed in Section 4 of this chapter.

3.2 A NOTE ON FUNCTIONAL FORM

A further need, identified above, in operationalising the utility based model of destination choice/shopping expenditure is to specify a form for the conditional indirect utility functions. Ideally what is required is a simple, flexible functional form that produces valid indirect utility functions and does not present unreasonable estimation problems. As will be shown, in practice it is difficult to concurrently meet all these features, necessitating trade-offs to be made.

Diewert (1974) has shown that indirect utility functions possess the following properties:

(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.

The selected functional form upon estimation should conform to these conditions. If these conditions are not met it will be clear that the function estimated is not an indirect utility function.

Adherence to these conditions, however, does not mean that the function estimated represents, or even approximately represents, the true indirect utility function implicitly used by the average consumer under study. Indeed certain frequently used functional forms, such as the linear expenditure system (Stone 1954) and the Cobb-Douglas utility function, display intuitively implausible and unduly restrictive properties. For example, the simple repackaging quality enhanced indirect utility function associated with a Cobb-Douglas direct utility function has the general form:

$$V = \prod_i \left(\frac{p_i}{\psi_i Y} \right)^{\alpha_i}$$

which yields demand functions:

$$C_i = \frac{\alpha_i Y}{p_i}$$

Not only is the demand for good C_i directly unaffected by the price of other goods but also by the value of its quality index, ψ_i.

An appealing family of functional forms that can serve as local first or second order approximations for any arbitrary utility function have become known as 'flexible functional forms'. Two members of this family are the 'indirect translog model' (Christensen, Jorgenson and Lau

1975) and the 'almost ideal demand system' (AIDS) (Deaton and Muellbauer 1980b).

Ignoring for the moment travel times and costs and the error terms, an AIDS version of the conditional indirect utility function shown in equation (3.20) can be expressed as:

$$\overline{V}_i = \left[\log Y - \alpha_0 - \alpha_1 \log p_i + \alpha_1 \Sigma \gamma_k b_{ik}\right] \times \left[\exp \sum_k \gamma_k b_{ik}\right] \frac{1}{\alpha_2} p_i^{-\alpha_3} \qquad (3.22)$$

and yields a conditional demand equation of the form:

$$g_i = \frac{Y}{p_i}\Big[\alpha_1 - \alpha_3 \alpha_0 - \alpha_3 \alpha_1 \log p_i + \alpha_3 \alpha_1 \sum \gamma_k b_{ik} + \alpha_3 \log Y\Big] \qquad (3.23)$$

A similar version of the translog model is:

$$\begin{aligned}\overline{V}_i = {} & \alpha_o - \alpha_1 \log p_1 - \alpha_1 \log \psi_i + (\alpha_1 + \alpha_2) \log Y - \frac{\alpha_3}{2} (\log p_i)^2 \\ & + \alpha_3 \log \psi_i \log p_i + (\alpha_3 - \alpha_4) \log Y \log p_i - {}^{\alpha_3}/_2 (\log p_i)^2 \\ & - (\alpha_3 - \alpha_4) \log \psi_i \log Y - (\alpha_3 + \alpha_4 + \alpha_5)(\log Y)^2\end{aligned} \qquad (3.24)$$

with the associated conditional demand function:

$$g_i = \frac{Y}{p_i} \frac{-\alpha_1 - \alpha_3 \log p_i + \alpha_3 \log \psi_i + (\alpha_3 - \alpha_4) \log Y}{-(\alpha_1 + \alpha_2) - (\alpha_3 - \alpha_4) \log p_i + (\alpha_3 - \alpha_4) \log \psi_i + (\alpha_3 + \alpha_4 + \alpha_5) \log Y} \qquad (3.25)$$

where ψ_i is given by equation (3.12). As can be seen, for the AIDS the indirect utility function tends to be quite complicated, making estimation difficult. The AIDS demand function, however, in budget share form is simple to estimate, i.e., with the dependent variable specified as $(p_i \, g_i)/Y$. The converse tends to be true for the translog model. In fact it seems generally to be true that for flexible functional forms estimation is difficult either for the demand equations or the indirect utility function or both. Furthermore, the apparent generality offered by flexible functional forms may be somewhat illusory. In particular the approximation to the true indirect utility function can only be accurate in the locality of specific (quality adjusted) price/income values, not over an entire sampling range. The necessary non-constancy of prices, quality variables and incomes in samples used for estimation invalidates the interpretation of flexible functional forms as standing in for any utility function.

In view of these drawbacks with the use of flexible functional forms an alternative, increasingly popular, approach has been to start with an easily estimatable demand function and derive (by differential equations and Roy's identity) the simplest form of indirect utility function compatible with it. For the shopping problem under study it is convenient to specify the demand model as linear-in-expenditure or log-linear-in-expenditure.

A linear-in-expenditure model corresponding to equation (3.21) (again momentarily neglecting the travel related variables and error terms) can generally be written as:

$$p_i \, g_i = \alpha_1 Y + f_1 (\log (p_i/\psi_i)) + \overline{\alpha}_0 \qquad (3.26)$$

where f_1 is linear-in-parameters and $\overline{\alpha}_0$ is a constant. A specific form of (3.26) is (using equation (3.12)):

$$p_i \, g_i = \alpha_1 Y - \alpha_2 \log p_i + \alpha_2 \sum \gamma_k b_{ik} + \alpha_3 \qquad (3.27)$$

A general form of $\overline{V}_i$ compatible with the linear-in-expenditure conditional demand model (3.26) is:

$$\overline{V}_i = f_2\left\{[f_1(\log(p_i/\psi_i)) + \alpha_1 Y]\exp[-\alpha_1 \log(p_i/\psi_i)]\right\} \tag{3.28}$$

where f_2 is any function, with a specific form of equation (3.28) being:

$$\overline{V}_i = \log\left[\alpha_3 - \alpha_2/\alpha_1 - \alpha_2 \log p_i + \alpha_2 \sum \gamma_k \ {}_{ik} + \alpha_1 Y\right]$$
$$- \alpha_1 \log p_1 + \alpha_1 \sum \gamma_k \ {}_{ik} \tag{3.29}$$

A similar exercise can be undertaken for a log-linear-in-expenditure model. Here the conditional demand function can generally be specified as:

$$\log (p_i g_i) = \alpha_1 \log Y + f_1 (\log (p_i/\psi_i)) + \overline{\alpha}_0 \tag{3.30}$$

of which a specific form is:

$$\log (p_i g_i) = \alpha_1 \log Y - \alpha_2 \log p_i + \alpha_2 \sum_k \gamma_k b_{ik} + \alpha_3 \tag{3.31}$$

A general form of conditional indirect utility function for the conditional demand function (3.30) is:

$$\overline{V}_i = f_2\left\{\frac{1}{\alpha_1 - 1} Y^{(1-\alpha_1)} + \frac{1}{\tilde{\alpha}_0} \exp\left[f_1[\log(p_i/\psi_i)]\right]\right\} \alpha_1 \neq 1,\ \tilde{\alpha}_0 \neq 0 \tag{3.32}$$

where $\tilde{\alpha}_0$ is an additional constant.

A useful specific form of (3.32) which, upon application of Roy's identity, yields (3.31) is:

$$\overline{V}_i = \frac{1}{\alpha_1 - 1} Y^{(1-\alpha_1)} - \frac{1}{\alpha_2} \exp(\alpha_3 - \alpha_2 \log p_i + \alpha_2 \log \psi_i) \tag{3.33}$$

Equation systems (3.22) and (3.24) and (3.30) and (3.32) form the basis of much of the work in this thesis. It should be realised that the Diewert restrictions if not naturally met, must be imposed on equations (3.28) and (3.32). Fuller discussion of the Diewert restrictions is reserved for Chapter 7.

4. A MODIFIED HANEMANN MODEL OF SHOPPING DESTINATION AND EXPENDITURE CHOICES WITH ACCESSIBILITY TREATED AS A QUALITY VARIABLE

In the previous subsection the embodiment of error terms and travel times and costs in the conditional indirect utility functions was ignored. One method of including these error terms and shopping destination accessibility, measured by travel times and costs, is as quality variables. The conditional indirect utility function of equation (3.20) can then be written as:

$$V_i = V^* \left(p_i/\overline{\psi}_i,\ Y\right) \tag{3.34}$$

where $\overline{\psi}_i = \overline{\psi}_i \left(B_i,\ t_i,\ c_i,\ \epsilon_i,\ \gamma,\ \gamma_c,\ \gamma_t\right)$

$$= \exp \left(\sum_k \gamma_k\, b_{ik} + \gamma_c\, c_i + \gamma_t\, t_i + \epsilon_i\right) \tag{3.35}$$

which represents a modified application of the discrete/continuous choice model system considered by Hanemann (1984). The parameter γ_c may be interpreted as representing the marginal utility of travel cost. Similarly, γ_t reflects the marginal utility of travel time.

Construct I as a polychotomous variable with values 1 to N and I = j if $g_j > 0$. Then, using an amended version of equation (3.20) to reflect the different treatment of travel times and costs:

$$\text{Prob}\left\{I = j\right\} = \text{Prob}\left\{V^* \, (p_j/\overline{\psi}_j, \, Y) \geq V^* \, (p_i/\overline{\psi}_i, \, Y), \right.$$
$$\left. i = 1, 2, \ldots, N, \, i \neq j\right\}$$
$$= \text{Prob}\left\{V^* \left(\frac{p_j}{\overline{\psi}_j Y}\right) \geq V \left(\frac{p_i}{\overline{\psi}_i Y}\right), \right.$$
$$\left. i = 1, 2, \ldots, N, \, i \neq j\right\} \quad (3.36)$$

Note that since indirect utility functions are homogeneous of degree zero in prices and income, no particular restriction is implied using ratios of prices to income rather than the variables themselves.

For the conditional indirect utility function $V_i = V\left(\frac{p_i}{\overline{\psi}_i Y}\right)$, i = 1, 2, ..., N, to be consistent with utility maximisation they must be decreasing in $p_i/\overline{\psi}_i$. Also in comparing any two conditional indirect utility functions Y remains constant and therefore can be eliminated. Equation (3.36) can, as a result, be simplified to:

$$\text{Prob}\left\{I=j\right\} = \text{Prob}\left\{p_j/\overline{\psi}_j \leq p_i/\overline{\psi}_i, \, i = 1, 2, \ldots, N, \, i \neq j\right\}$$
$$= \text{Prob}\left\{\log\overline{\psi}_j - \log p_j \geq \log\overline{\psi}_i - \log p_i, \, i = 1, 2, \ldots, N, \, i \neq j\right\}$$
$$= \text{Prob}\left\{\epsilon_j + \gamma_c c_j + \gamma_t t_j + \sum_k \gamma_k b_{jk} - \log p_j \geq \right.$$
$$\epsilon_i + \gamma_c c_i + \gamma_t t_i + \sum_k \gamma_k b_{ik} - \log p_i,$$
$$\left. i = 1, 2, \ldots, N, \, i \neq j\right\} \quad (3.37)$$

If the ϵ_is are independently and identically distributed (iid) with an extreme value type 1 distribution then the resulting form of Prob $\left\{I = j\right\}$ is a logit model,

$$\text{Prob}\{I=j\} = P_j = \frac{\exp\left[\frac{1}{\mu}\sum_k \gamma_k b_{jk} + \frac{\gamma_c}{\mu} c_j + \frac{\gamma_t}{\mu} t_j + \frac{1}{\mu}\log p_j\right]}{\sum_i \exp\left[\frac{1}{\mu}\sum_k \gamma_k b_{ik} + \frac{\gamma_c}{\mu} c_i + \frac{\gamma_t}{\mu} t_i + \frac{1}{\mu}\log p_i\right]} \tag{3.38}$$

where μ is the logistic scale parameter.

While the choice of shopping destination (= Prob $\{g_i > 0\}$), the discrete element of the utility maximising problem, can be determined simply by evaluating $p_i/\overline{\psi}_i$, to obtain the amount of shopping goods purchased (i.e., the value of $p_j g_j$) it is necessary to specify a form for the conditional indirect utility function. Suppose V_j is of the form shown in equation (3.33) with ψ_j replaced by $\overline{\psi}_j$. Then, as noted earlier, the associated demand function can be derived through application of Roy's identity.*

---

* The discrete choice model could have also been directly derived from the indirect utility function. The conditional function is:

$$V_j = \frac{1}{\alpha_1 - 1} Y^{(1-\alpha_1)} - \frac{1}{\alpha_2}\exp(\alpha_3 - \alpha_2 \log p_j + \alpha_2 \log \overline{\psi}_j) \tag{3.33 rptd}$$

In comparing V_j with V_i the Y term is eliminated so:

$$\text{Prob}\{V_j \geq V_i\} = \text{Prob}\{\tfrac{1}{\alpha_2}\exp(\alpha_3 - \alpha_2 \log p_j + \alpha_2 \log \overline{\psi}_j) \geq \tfrac{1}{\alpha_2}\exp(\alpha_3 - \alpha_2 \log p_i + \alpha_2 \log \overline{\psi}_i);\ i = 1, 2, \ldots, N,\ i \neq j\} \tag{3.1F}$$

Taking logarithms and simplifying (3.1F) yields:

$$\text{Prob}\{V_j \geq V_i\} = \text{Prob}\{\log \overline{\psi}_j - \log p_j \geq \log \overline{\psi}_i - \log p_i;\ i = 1, 2, \ldots, N,\ i \neq j\}$$

or,

$$\text{Prob}\{I = j\} = \text{Prob}\{\epsilon_j + \gamma_c c_j + \gamma_t t_j + \sum_k \gamma_k b_{jk} - \log p_j \geq \epsilon_i + \gamma_c c_i + \gamma_t t_i + \sum_k \gamma_k b_{ik} - \log p_i,\ i = 1, 2, \ldots, N,\ i \neq j\}$$

which is equivalent to equation (3.37). However, since the derivation of (3.37) applies to all forms of V_j $(p_j/\overline{\psi}_j, Y)$ it is the more general.

For shopping goods purchased at destination j the demand function is:

$$g_j = -\frac{\partial V_j / \partial p_j}{\partial V_j / \partial Y} = \frac{Y^{\alpha_1}}{p_j} \exp(\alpha_3 - \alpha_2 \log p_j + \alpha_2 \log \overline{\psi}_j)$$

(3.39)

Equation (3.39) can be inverted to obtain an expression for log ($\overline{\psi}_j$):

$$\log(\overline{\psi}_j) = -\frac{1}{\alpha_2} (\alpha_3 - (\alpha_2 + 1) \log p_j + \alpha_1 \log Y - \log g_j) \quad (3.40)$$

By denoting $\overline{\lambda}_j$ as $\sum_k \gamma_k b_{jk} + \gamma_c c_j + \gamma_t t_j - \log p_j$, that is as the non-stochastic elements of equation (3.37), then:

$$\overline{\psi}_j = \exp(\overline{\lambda}_j + \log p_j + \epsilon_j)$$

and from (3.40):

$$-\epsilon_j = \frac{1}{\alpha_2} (\alpha_3 - (\alpha_2 + 1) \log p_j + \alpha_1 \log Y - \log g_j) + \overline{\lambda}_j + \log p_j \quad (3.41)$$

To obtain the conditional probability distribution for the set error terms $\{\epsilon \mid I = j\}$ recall that an extreme value type 1 distribution was used to derive the logit model of equation (3.38). The distribution has the form:

$$F_\epsilon^j = \exp(-r/\mu)\left[\exp - \sum_i \exp(-\overline{\lambda}_j/\mu + \overline{\lambda}_i/\mu + r/\mu)\right]/\mu$$

(3.42)

The conditional marginal density for f $_{\epsilon_j \mid I = j}$ is therefore:

$$f_{\epsilon_j|I=j}(\epsilon_j) = F^j/P_j$$

$$= \sum_i \exp(-\overline{\lambda}_j/\mu + \overline{\lambda}_i/\mu)\exp(-\epsilon_j/\mu)$$

$$\times \exp\Big[-\sum_i \exp(-\overline{\lambda}_j/\mu + \overline{\lambda}_i/\mu$$

$$+ \epsilon_j/\mu)\Big]^{\mu} \qquad (3.43)$$

Substituting for ϵ_j the RHS of equation (3.41) gives:

$$f_{g_j|I=j}(g_j) = \sum_i \exp(-\overline{\lambda}_j/\mu + \overline{\lambda}_i/\mu)\exp\Big[(\alpha_2\mu)^{-1}(\alpha_3 -$$

$$(\alpha_2 + 1)\log p_j + \alpha_2 Y - \log g_j) + \overline{\lambda}_j/\mu$$

$$+ \frac{1}{\mu}\log p_j\Big]\exp\Big(-\sum_i \exp(-\overline{\lambda}_j/\mu + \overline{\lambda}_i/\mu)$$

$$\times \exp\Big[(\alpha_2\mu)^{-1}(\alpha_3 - (\alpha_2 + 1)\log p_j$$

$$+ \alpha_2 Y - \log g_j) + \overline{\lambda}_j/\mu + \frac{1}{\mu}\log p_j\Big]\Big)/\mu$$

$$= \sum_i \exp(\overline{\lambda}_i/\mu)\exp[\alpha_3(\alpha_2\mu)^{-1}]p_j^{-1/\alpha_2\mu}$$

$$\times Y^{\alpha_1/\alpha_2\mu} g_j^{-1/\alpha_2\mu} \exp\Big[\sum_i \exp(\overline{\lambda}_i/\mu)$$

$$\times \exp[\alpha_3(\alpha_2\mu)^{-1}]p_j^{-1/\alpha_2\mu} Y^{\alpha_1/\mu\alpha_2}$$

$$\times g_j^{-1/\alpha_2\mu}\Big]^{\mu^{-1}}$$

(3.44)

The discrete/continuous system of equations (3.38) and (3.39) can now be estimated using full information maximum likelihood. The likelihood function for a sample of Q individuals is given by:

$$\Gamma^* = \prod_{q=1}^{Q}\Big(f_{g_j*q|I=j}(g_q^*) \times P_{j*q}\Big) \qquad (3.45)$$

where P^*_{jq} is given by equation (3.38) and $f_{g_j{}^*q|I=j}$ by equation (3.44).

Evaluation of this likelihood function in principle will provide consistent and asymptotically efficient estimates for all unknown parameters. The likelihood function, however, may possess multiple roots and convergence to the global maximum is as a consequence not assured.

Fortunately, an alternative two stage estimator is available. The logit model of equation (3.38) can be estimated by maximum likelihood methods as described for instance, by McFadden (1974). The log likelihood function is:

$$\Gamma = \sum_{q=1}^{Q} \sum_{i=1}^{N} \left(k_{iq} \log (P_{iq}) \right) \tag{3.46}$$

where $k_{iq} = 1$ if alternative i is chosen and 0 otherwise. P_{iq} is given by the RHS of equation (3.38). Standard computer packages are available to evaluate this function and provide estimates of γ, γ_c, γ_t and μ.

To obtain an estimator for the continuous choice model, rearrange equation (3.41):

$$\exp(-\epsilon_j) = \exp\left(\frac{\alpha_3}{\alpha_2} - \log p_j - \frac{1}{\alpha_2} \log p_j + \frac{\alpha_1}{\alpha_2} \log Y - \frac{1}{\alpha_2} \log g_j\right) \exp(\overline{\lambda}_j)\, p_j$$

$$(p_j g_j)^{1/\alpha_2} = \exp(\alpha_3/\alpha_2) Y^{\alpha_1/\alpha_2} \exp(\overline{\lambda}_j + \epsilon_j)$$

$$\log (p_j g_j) = \alpha_3 + \alpha_1 \log Y + \alpha_2 (\overline{\lambda}_j + \epsilon_j) \tag{3.47}$$

where $(pg)_j = p_j\, g_j$. The expected value of $\log (pg)_j$ is:

$$E\{\log (pg)_j | I = j\} = \alpha_3 + \alpha_1 \log Y + \alpha_2 (\overline{\lambda}_j + E\{\epsilon_j | I = j\}) \tag{3.48}$$

but (see, for example, Johnson and Kotz 1970 p. 278),

$$E\{\epsilon_j | I = j\} = \int_{-\infty}^{+\infty} \epsilon_j \sum_i \exp(- \overline{\lambda}_j/\mu + \overline{\lambda}_i/\mu)$$

$$\times \exp(- \epsilon_j/\mu) \exp\left[- \sum_i \exp(- \overline{\lambda}_j/\mu + \overline{\lambda}_i/\mu - \epsilon_j/\mu) \right] \mu^{-1} d\epsilon_j$$

$$= \mu \left(\log \sum_i \exp(- \overline{\lambda}_j/\mu + \overline{\lambda}_i/\mu) + \zeta \right) \quad (3.49)$$

where ζ is Eulers constant (≈ 0.577). Therefore:

$$E\{\log (pg)_j | I = j\} = \alpha_3 + \alpha_1 \log Y + \alpha_2 \mu \times \left(\log \sum_i \exp(\overline{\lambda}_i/\mu) + 0.577 \right) \quad (3.50)$$

Equation (3.50) pertains to an individual. When an aggregation of individuals is being considered the appropriate model is:

$$\log (pg)^*_{j*q} = \alpha_3 - \alpha_1 \log Y_q + \alpha_2 \mu \times \left(\log \sum_i \exp(\overline{\lambda}_{iq}/\mu) + 0.577 \right) + v_q \quad (3.51)$$

where v_q is an individual specific error term that is distributed iid extreme value type 1 with $E(v_q) = 0$. The estimated values of $\overline{\lambda}_{iq}$ and μ obtained from application of equation (3.46) to equation (3.38) are used to form the term $\left(\mu \log \sum_i \exp(\overline{\lambda}_{iq}/\mu) + \mu\, 0.577 \right)$.

The shopping expenditure model to be estimated is then:

$$\log (pg)^{*}_{j*q} = \alpha_3 + \alpha_1 \log Y + \alpha_2 \hat{\mu} \times \left(\log \sum_i \exp (\hat{\lambda}_{iq} / \hat{\mu} + 0.577 \right) + \overline{v}_q \tag{3.52a}$$

where

$$\overline{v}_q = v_q + \alpha_2 \left(\hat{\mu} \log \sum_i \exp (\hat{\lambda}_{iq} / \hat{\mu}) - \mu \log \sum_i (\overline{\lambda}_{iq} / \mu) + (\hat{\mu} - \mu) \, 0.577 \right) \tag{3.52b}$$

Equation (3.52a) can be estimated by OLS to obtain consistent, but inefficient, estimates of α_1, α_2 and α_3. The conventionally calculated variance estimates associated with these parameters will, moreover, be biased because the $\overline{v}_q$ are non normally distributed and from (3.52b), heteroskedastic. To obtain unbiased variance estimates the consistently estimated values for $\hat{\gamma}$, $\hat{\gamma}_c$, $\hat{\gamma}_t$, $\hat{\mu}$, $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ from equations (3.38) and (3.51) should be used as starting values in a one step Newton - Raphson evaluation of the likelihood function (3.45).* The parameter estimates obtained will be consistent, asymptotically normal and efficient. The unbiased estimated variance-covariance matrix can be extracted in the usual way by reference to the information matrix associated with (3.45).

---

* At least in some contexts, however, it would appear that a one step evaluation is inadequate. For instance, Berkovic and Rust in estimating a nested logit model of automobile holdings using full information maximum likelihood, following sequential estimation of their model system, report: 'In theory, one Newton-step should yield efficient parameter estimates that are close to the parameters that maximise the full likelihood function. In practice, it appears that the parameter estimates obtained in this manner are not stable and may be relatively far away from the parameter values that maximise the full likelihood function. It appears that in order to guarantee convergence estimates, one must iterate until stationarity of the likelihood function is obtained rather than perform only one step' (Berkovic and Rust 1985, p. 279). In a similar context, this has also been the experience of Hensher (1986).

5. A FULL CONDITIONAL INDIRECT UTILITY MODEL

The shopping application of the modified Hanemann model as presented above, although highlighting the relationship between the error terms in the utility model and the demand model, would seem to suffer from two major disadvantages.

The first disadvantage is that the sole source of error in the demand model **for an individual** is assumed to derive from uncertainty concerning the quality index (ψ_j, I = j). No error is introduced into the demand model as a result, for instance, of slight misspecification of the indirect utility function or (incorrect) omission of quality variables associated with the Hicksian composite commodity. Further, it is assumed that the budget constraint is observed with certainty. In most applications this condition is unlikely to be met. For the shopping problem under study the p_is refer to price indices at shopping destinations. This aggregate concept ignores many individual variations such as differences in purchasing patterns within a shopping destination.

A second, possibly more telling, disadvantage with the Hanemann model when applied to the shopping destination/expenditure choice problem has to do with the treatment of travel time and costs. In particular treating destination accessibility as a quality variable, although convenient in terms of producing tractable model forms, is somewhat removed from the spirit of theory presented in Section 2 of this chapter which introduced travel time and costs as constraints on choice. Once travel times and costs are included as constraints, however, the choice of shopping destination (= Prob $\{ g_j > 0 \}$) cannot simply be evaluated by reference to the p_i / ψ_is but must also take account of the variation in remaining income ($Y_i = Y - c_i$) and remaining time ($T_i = T - t_i$).

This drawback to applying the Hanemann model is not merely restricted to the current shopping destination expenditure choice environment. It is common in situations with an element of discrete choice for there to be some fixed costs. In applied demand studies such costs, which are invariant to degree of use, have been traditionally incorporated by subtracting them from income. The conditional indirect utility function then becomes:

$$V_i = V_i \left(\frac{p_i}{\psi_i Y_i}, \frac{1}{Y_i}\right)$$

where 1 represents the normalised price of the Hicksian composite good and Y_i income less fixed costs K_i. It can clearly be seen that the Hanemann method is unserviceable when the conditional indirect utility function is of this form.

To remain in harmony with the theory presented in Section 2 it is necessary to evaluate the full conditional indirect utility function in predicting the discrete choice of shopping destination. Virtually all previous work in this small but expanding field of discrete/continuous modelling based on an integrated economic consumer theoretic framework has directly used the conditional indirect utility function in the discrete choice model. This work includes published articles by Trost and Lee (1984), Hill (1983) and Dubin and McFadden (1984), reporting empirical research where the discrete choice is polychotomous. Four further unpublished pieces are Hay (1979), Mannering and Winston (1985), Hensher and Milthorpe (1985) and Train and Lohrer (1983).* A disadvantage of using the full conditional indirect function is that forms which produce interesting demand functions are inevitably non-linear. This makes the discrete choice model slightly more difficult to estimate. **

---

* In addition to these works and the work of Hanemann, the theoretical relationship between discrete and continuous choices has been explored, in the context of welfare evaluation, by Small and Rosen (1981).

** Of the authors just mentioned, all empirical studies except Dubin and McFadden, Mannering and Winston, assumed the indirect utility function to be linear and did not directly derive demand functions from the utility function. The link between much of this work and theory is therefore rather tenuous.

The form for the indirect utility function used in presenting the modified Hanemann model, namely equation (3.33), is reasonably non-linear once Y is replaced by Y_i, making estimation quite difficult. An easier form to estimate is given by equation (3.29), appropriately modified to incorporate remaining time and income terms:

$$\overline{V}_j = \log\left(\alpha_3 - \frac{\alpha_2}{\alpha_1} - \alpha_2 \log \psi_j + \alpha_1 (Y - c_j) + \alpha_4 (T - t_j)\right) - \alpha_1 \log p_j + \alpha_1 \log \psi_j \quad (3.53)$$

It can be seen that (3.53) is a special case of the general conditional indirect utility function arising from the theoretical model specified in equations (3.1) - (3.3). This conditional indirect utility function was originally specified in equation (3.20) as $V_j = V^* (p_j/\psi_j, t_j, c_j, Y, \epsilon)$, but can more precisely be written as: $V_j = V^* (p_j/\psi_j, T - t_j, Y - c_j, \epsilon)$.

An error term ϵ_j can be added to equation (3.53) to account for the uncertainty in observing individually obtained utility. The the probability that shopping destination j is chosen, P_j, is:

$$\begin{aligned} P_j &= \text{Prob}\{I = j\} = \text{Prob}\{\overline{V}_j + \epsilon_j \geq \overline{V}_i + \epsilon_i, i = 1, 2, .., N, i \neq j\} \\ &= \text{Prob}\{\log[\alpha_3 - \frac{\alpha_2}{\alpha_1} - \alpha_2 \log p_j + \alpha_2 \sum_k \gamma_k b_{jk} + \alpha_1 (Y - c_j) \\ &\quad + \alpha_4 (T - t_j)] - \alpha_1 \log p_j + \alpha_1 \sum_k \gamma_k b_{jk} + \epsilon_j \\ &\quad \geq \log[\alpha_3 - \frac{\alpha_2}{\alpha_1} - \alpha_2 \log p_i + \alpha_2 \sum_k \gamma_k b_{ik} + \alpha_1 (Y - c_i) \\ &\quad + \alpha_4 (T - t_i)] - \alpha_1 \log p_i + \alpha_1 \sum_k \gamma_k b_{ik} + \epsilon_i\} \end{aligned} \quad (3.54)$$

When the ϵ_i are iid extreme value type 1 the choice probabilities are given by:

$$P_j = \frac{\exp\{\alpha_1 \sum_k \gamma_k b_{jk} - \alpha_1 \log p_j + \log[\alpha_3 - \frac{\alpha_2}{\alpha_1} - \alpha_2 \log p_j + \alpha_2 \sum_k \gamma_k b_{jk} + \alpha_1(Y - c_j) + \alpha_4 (T - t_j)]\}}{\sum_i \exp\{\alpha_1 \sum_k \gamma_k b_{ik} - \alpha_1 \log p_i + \log[\alpha_3 - \frac{\alpha_2}{\alpha_1} \log p_i + \alpha_2 \sum_k \gamma_k b_{ik} + \alpha_1(Y - c_i) + \alpha_4(T - t_i)]\}} \quad (3.55)$$

The logit model of equation (3.55) is not dissimilar to the type of logit model applied when the alternatives used in estimation are not elemental alternatives, but rather represent amalgamations of elemental alternatives. This situation was discussed in Chapter 2, Section 3. From equation (2.5) with $\theta = 1$ and $\alpha_{j\ell} = \alpha_{i\ell}$, the correct form of logit model for the aggregate alternatives when $\overline{V}_i$ is linear for the elemental alternatives is:

$$P_j = \frac{\exp\left[\sum_{\ell=1}^{L} \alpha_\ell z_{\ell j} + \log\left(\sum_{\ell=L+1}^{M} \alpha_\ell s_{\ell j}\right)\right]}{\sum_i \exp\left[\sum_{\ell=1}^{L} \alpha_\ell z_{\ell i} + \log\left(\sum_{\ell=L+1}^{M} \alpha_\ell s_{\ell i}\right)\right]} \quad (3.56)$$

Unfortunately, as noted in Chapter 2, the parameters $\alpha_{L+1}, \alpha_{L+2}, \ldots, \alpha_M$ are only identifiable up to a multiplicative constant. For the particular problem under study, this restriction may be properly circumvented, because of relationships between parameters within the log term and other parameters in $\overline{V}_i$. It does, however, complicate the estimation process.

Another variant of equation (3.28) is:

$$\overline{V}_i = \left(\alpha_3 - \alpha_2 \log(p_i/\psi_i) + \alpha_1 (Y - c_i) + \alpha_4 (T - t_i)\right) \exp\left(-\alpha_5 \log(p_i/\psi_i)\right) \quad (3.57)$$

With the addition of iid error terms taking an extreme value type 1 distribution to (3.57), the shopping destination choice probabilities can be described by:

$$P_j = \frac{\exp\left[\left[\alpha_3 - \alpha_2 \log(p_j/\psi_j) + \alpha_1(Y-c_j) + \alpha_4(T-t_j)\right](p_j/\psi_j)^{-\alpha_5}\right]}{\sum_i \exp\left[\left[\alpha_3 - \alpha_2 \log(p_i/\psi_i) + \alpha_1(Y-c_i) + \alpha_4(T-t\ )\right](p_i/\psi_i)^{-\alpha_5}\right]} \tag{3.58}$$

The MNL model of (3.58) can be estimated using maximum likelihood. The log likelihood function for a sample of Q individuals is:

$$\Gamma = \sum_q \sum_i k_{iq} \log (P_{iq}) \tag{3.59}$$

where P_{iq} is given by the RHS of equation (3.58) appropriately modified to take into account different individuals.

The shopping expenditure model associated with equation (3.57) is:

$$(pq)_{iq} = \frac{\alpha_5 \alpha_3}{\alpha_1} + \frac{\alpha_2}{\alpha_1} - \frac{\alpha_2 \alpha_5}{\alpha_1} \log (p_{iq}/\psi_{iq})$$
$$+ \alpha_5 (Y_q - c_{iq}) + \frac{\alpha_5 \alpha_4}{\alpha_1} (T - t_{iq}) + u_{iq} \tag{3.60}$$

where u_{iq} is an additive error term. For reasons well emphasised by the Hanemann model, in estimating the discrete/continuous choice model system of equations (3.55) and (3.56) the dependency of error terms in the two models needs to be recognised. This recognition, however, will now take a statistical form rather than be theoretically derived. Two methods of statistically accounting for this dependency are discussed in Chapter 7.

6. CONCLUSION

In this chapter a theoretical framework has been established for analysing shopping behaviour, particularly, shopping destination choice. This theory also provides a link between shopping destination choices and decisions regarding the intensity of participation in shopping activities, measured by the level of shopping expenditure. The theoretical framework was built around the paradigm of economic utility maximisation and from this base a number of empirically estimable models were derived. These models clearly demonstrated the relationship between shopping destination and expenditure choices.

The remainder of this thesis is devoted to empirically delving into a number of aspects of the theoretical framework presented above. In the next three chapters aspects of the destination choice decision are empirically analysed. Chapter 4 contains some basic mode/destination choice models for categories of food shopping, Chapter 5 investigates specification of the destination choice set for major household grocery shopping and Chapter 6 the linking of destination choice decisions between the various categories of food shopping. These chapters, in common with virtually all past research, utilise a $\overline{V}_i$ specification which is linear in the parameters and the variables. The linear specification may be viewed as a first order approximation to the true, non-linear, form. In Chapter 7 a non-linear form, similar to that shown in equation (3.57) is utilised for $\overline{V}_i$ and an integrated shopping destination and expenditure choice model estimated, the application context being urban grocery shopping. It will be seen from the empirical estimates that for the data set used in this study, for grocery shopping, the linear form for $\overline{V}_i$ may be a reasonable approximation. The data set is described in the next chapter.

Note from the above that in order to impregnate the empirical study with a high behavioural context, shopping destinations have been identified at a high level of spatial specificity, in particular, by individual stores. (In fact, in Chapter 6, destinations are identified by sections of individual stores.) Care has been taken, however, in this chapter to identify the i = 1, 2, ..., N, generically as 'shopping destinations'. Provided appropriate price indices can be formed there is no bar to applying the theory developed in this chapter to shopping centres (as may be required for traffic generation work) or shopping zones (as may be required in transport studies).

APPENDIX 3A

PROOF OF ROY'S IDENTITY WHEN APPLIED TO A QUALITY ENHANCED CONDITIONAL RANDOM INDIRECT UTILITY FUNCTION

Begin with a conditional maximisation problem similar to that defined in equation system (3.6a) - (3.6c):

$$\max U_i = U_i\ (g_i, B_i, \overline{Z}, L, \epsilon) \tag{3.A1a}$$

$$\text{subject to: } Y = p_i g_i + \overline{Z} + c_i \tag{3.A1b}$$

$$L = T - t_i \tag{3.A1c}$$

The maximisation problem can be solved by forming the Lagrangian,

$$G = U_i\ (g_i, \overline{Z}, T - t_i, B_i, \epsilon) + \lambda(Y - p_i g_i - \overline{Z} - c_i) \tag{3.A2}$$

and obtaining first-order conditions:

$$\frac{\partial U_i}{\partial g_i} = \lambda p_i \tag{3.A3a}$$

$$\frac{\partial U_i}{\partial \overline{Z}_i} = \lambda \tag{3.A3b}$$

$$Y - p_i g_i - \overline{Z} - c_i = 0 \tag{3.A3c}$$

The marginal rate of substitution between consumption of groceries and the Hicksian composite commodity is:

$$\frac{\partial U_i / \partial g_i}{\partial U_i / \partial \overline{Z}_i} = p_i \tag{3.A4}$$

and the conditional grocery demand function resulting from solving (3.A1a) - (3.A1c) is:

$$g_i = g_i \ (p_i, B_i, t_i, c_i, Y, \epsilon) \tag{3.A5}$$

The conditional indirect utility function can now be formed as:

$$V_i(p_i, B_i, t_i, c_i, Y, \epsilon) = U_i[g_i(p_i, B_i, t_i, c_i, Y, \epsilon), \bar{Z}_i(p_i, B_i, t_i, c_i, Y, \epsilon)] \tag{3.A6}$$

Partially differentiating (3.A6) w.r.t. p_i, then:

$$\frac{\partial V_i}{\partial p_i} = \frac{\partial U_i}{\partial g_i}\frac{\partial g_i}{\partial p_i} + \frac{\partial U_i}{\partial \bar{Z}_i}\frac{\partial \bar{Z}_i}{\partial p_i} \tag{3.A7}$$

But, $\frac{\partial U_i}{\partial g_i} = \lambda p_i$. Therefore:

$$\frac{\partial V_i}{\partial p_i} = \lambda p_i \frac{\partial g_i}{\partial p_i} + \frac{\partial \bar{Z}_i}{\partial p_i} = \lambda \left(p_i \frac{\partial g_i}{\partial p_i} + \frac{\partial \bar{Z}_i}{\partial p_i}\right) \tag{3.A8}$$

The budget constraint must also be satisfied:

$$Y = p_i g_i \ (p_i, B_i, t_i, c_i, Y, \epsilon) + \bar{Z}_i \ (p_i, B_i, t_i, c_i, Y, \epsilon) + c_i \tag{3.A9}$$

Differentiating (3.A9) w.r.t. p_i gives:

$$g_i(p_i, B_i, t_i, c_i, Y, \epsilon) + \frac{\partial \bar{Z}_i}{\partial p_i} + p_i \frac{\partial g_i}{\partial p_i} = 0$$

or,

$$\frac{\partial \bar{Z}_i}{\partial p_i} + p_i \frac{\partial g_i}{\partial p_i} = -g_i(p_i, B_i, t_i, c_i, Y, \epsilon) \tag{3.A10}$$

Substituting (3.A10) into (3.A8) yields:

$$\frac{\partial V_i}{\partial p_i} = -\lambda \ g_i \ (p_i, B_i, t_i, c_i, Y, \epsilon) \tag{3.A11}$$

Next differentiate $V_i(.)$ w.r.t. Y:

$$\frac{\partial V_i}{\partial Y} = \frac{\partial U_i}{\partial g_i}\frac{\partial g_i}{\partial Y} + \frac{\partial U_i}{\partial \overline{Z}_i}\frac{\partial \overline{Z}_i}{\partial Y} \qquad (3.A12)$$

$$= \lambda\,(p_i \frac{\partial g_i}{\partial Y} + \frac{\partial \overline{Z}_i}{\partial Y})$$

Then, differentiating the budget constraint w.r.t. Y:

$$1 = p_i \frac{\partial g_i}{\partial Y} + \frac{\partial \overline{Z}_i}{\partial Y} \qquad (3.A13)$$

and substituting (3.A13) into (3.A12) gives:

$$\frac{\partial V_i}{\partial Y} = \lambda \qquad (3.A14)$$

Combining (3.A14) and (3.A11) gives:

$$- \frac{\partial V_i/\partial p_i}{\partial V_i/\partial Y} = g_i\ (p_i, B_i, t_i, c_i, Y, \epsilon) \qquad (3.A15)$$

which is Roy's identity.

CHAPTER 4

DATA DESCRIPTION AND SOME BASIC MODELS OF FOOD SHOPPING MODE/DESTINATION CHOICES

1. INTRODUCTION

Having specified a sound economic theoretic base for analysing shopping destination and expenditure choices it is now timely to empirically examine certain aspects of these decisions. The present chapter acts as an initiation to this task. It is sectioned into two parts. The initial portions are devoted to a discussion of the data base used in this study. The final portions of this chapter describe the estimation of some basic models of food shopping mode/destination choices.

2. DATA

The data used in this thesis were from the Adelaide Travel Demand and Time Allocation Survey (ATDATAS), collected by the author in October/November 1980 on behalf of the South Australian Department of Transport and Australian Road Research Board.

The overall objective of this survey was to collect a data set which would permit investigation of the travel decision making process and facilitate the development of more realistic models of travel behaviour. The primary mechanism for achieving this objective was to adopt the framework developed by human activity researchers in which travel is viewed explicitly as a derived demand, that is, as an outcome of the demand for other activities. This framework also places emphasis on the existence of constraints (both of a spatial and temporal nature) as affecting the amount of travel demanded by individuals.

In working towards a survey design three specific aims were initially enumerated from the overall objective and framework. These were:

(i) to collect information on all activities, not just the travel activity, in order that travel activities be analysed in conjunction with other activities,

(ii) to collect a data set which would permit comparison with and validation of the 1977 Metropolitan Adelaide Data Base Study (MADBS),

(iii) to collect a data set that would enable detailed modelling of destination and journey structure choices, as these decisions were seen as offering the most promising avenue for improvement to existing travel forecasting techniques, and

These aims suggested a two part survey design. The first part consisted of having selected households record all that they did, for a period of a week, in activity diaries. At the end of that period, diaries were collected and adult members from each household selected to participate in a personal interview. These interviews asked details of work and shopping trips undertaken by a randomly selected fully employed household member and the main household shopper, respectively. As well, information was sought on personal and household characteristics. The interviews involved between 45 and 60 minutes elapsed time.

The location of the survey was five local government areas (LGAs), Burnside, Kensington and Norwood, St. Peters, Payneham and Campbelltown, in the eastern and north-eastern suburbs of Adelaide (see Figure 4.1). These LGAs cover an area of approximately 60 sq. km. The eastern/lower north eastern region of Adelaide was chosen because of limited modal availability, thus simplifying application of

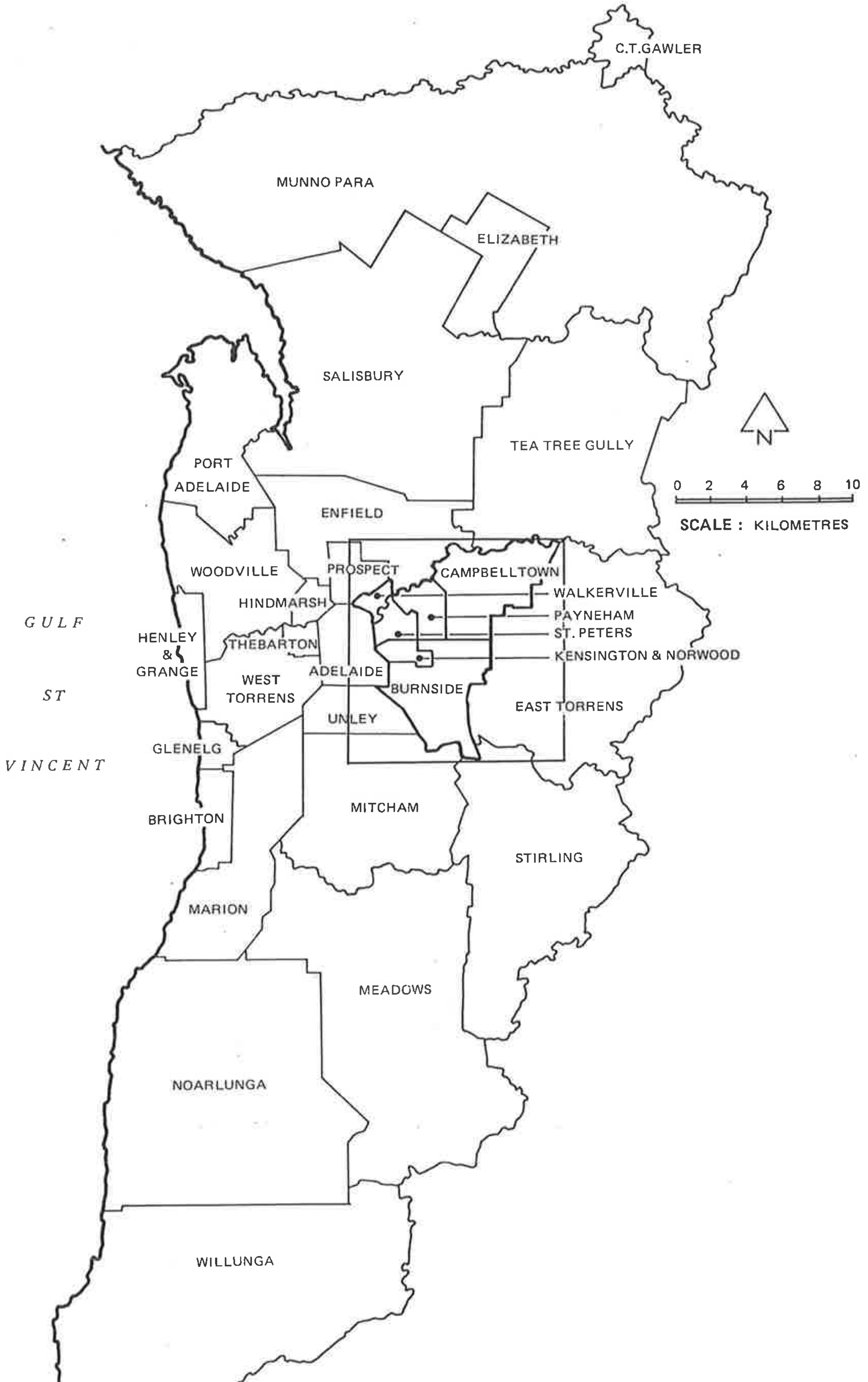


FIGURE 4.1: MAP OF STUDY AREA

exploratory techniques, and because major public transport infrastructure investment was planned for the area in the short to medium term.

The sampling scheme used was somewhat different from the norm. In 1977, a major transport study home interview survey was conducted throughout the Adelaide metropolitan area. The sample for this survey, for private dwellings, was randomly drawn from Electricity Trust of South Australia records at the rate of 1 in 56. In all, completed interviews were obtained for 4,440 dwellings, representing a response rate of 78%. For the five LGAs mentioned in the eastern and north eastern suburbs of Adelaide, 673 households participated in the 1977 survey, with no contact being made with 5.6% of households and 13.2% of households approached, refusing to complete the interview. Extensive checks were conducted on the MADBS home interview survey (HIS) data against census data, to ensure that a representative sample had been obtained. No substantial differences were found between the MADBS HIS sample and census data. Details of the MADBS HIS are given in Pak Poy and Associates (1978). The sampling scheme for ATDATAS then involved returning to 534 randomly selected households, residing in the five mentioned LGAs, who had participated in the 1977 survey. This sampling method resulted in a substantial pool of households participating in both the 1977 and 1980 surveys, but also a minority of households who had moved into these dwellings in the period 1977-1980 only participating in the 1980 survey. On the debit side some bias was undoubtably introduced into the 1980 survey as a result of this procedure because non-respondents from the 1977 survey were automatically excluded from the 1980 survey. It is likely, however, that this source of bias is of a minor nature.

First contact with households surveyed in 1980 was by post. The letter they received briefly informed them of the survey and requested their co-operation. Next, personal contact was made by the interviewer and diaries distributed and explained to every household member 12 years of age of older. Also an appointment was made to collect the diaries and conduct the interviews. In the middle of the recording period interviewers were instructed to again contact households (normally by telephone) to discuss any problems that may have occurred. At the end of the recording period diaries were collected at the appointed time and three sets of questions asked. One set sought information on the household's socio-economic characteristics and followed data normally collected in transportation study HISs. Another set was directed at a randomly selected employed household member and concerned mode choice for the journey to work. The third set requested information from the main household shopper on the household's food shopping arrangements. All uncooperative households were asked whether they were resident in the dwelling in 1977. Also an attempt was made to collect skeleton socio-economic and travel information. The shopping questionnaire is reproduced in Appendix 4A.

Concerning the shopping questionnaire, prior to the main survey a small telephone survey of 50 Adelaide shoppers was conducted. Shoppers were asked to supply information on the outlets usually chosen for meat, grocery and greengrocery shopping and their reasons for liking these outlets better than others, disregarding travel factors. This was a method for identifying destination attractiveness attributes reviewed in Chapter 2. For grocery shopping, selection of goods, prices and store convenience dominated reported choice attributes. For meat and greengrocery shopping, quality of merchandise, prices and store convenience were the most frequently mentioned attributes.

The main survey shopping questionnaire sought information for each category of food shopping on the frequency and timing of travel, the normal level of shopping expenditure, the shopping outlet usually patronised and method of travel to that outlet, and alternative outlets and modes considered by the individual.

The latter two sets of questions were designed to elicit respondents' perceived mode/destination choice sets for each food shopping category. The method was for the interviewer to fill out a table based on the respondent's answers. The respondent was first asked for the outlet normally used for that category of food shopping. Data pertaining to that outlet were then obtained; namely perceived attractiveness information and how going to that store was normally fitted into the respondent's schedule (i.e. the normal connecting activities). Next the respondent was asked for the normal method of travel to that outlet. After obtaining some information about the normal method of travel, other ways the respondent **could** travel to that outlet were explored and information collected about these. This completes the data gathered regarding the outlet normally used. Next the respondent was asked for another store that they would consider using for this category of food shopping if their normal store was unavailable. Perceived store attribute and modal information was collected in a similar manner to that for the normal store. This process was repeated until the set of alternative stores for the respondent was exhausted. An example of some of the output to emerge from this questioning is shown in Figure 4.2.

Overall information on food shopping patterns was also collected which included the relative location of shopping outlets, the temporal spacing of shopping activities and the overall food shopping travel pattern selected.

For each shopping outlet mentioned respondents were asked to rate the outlet in terms of the price, selection/ quality and convenience attributes. To do this they were provided with a five point rating scale with a value range from 'far above average' (5) to 'far below average' (1). In addition, for each mode, respondents were asked to supply information on travel times, travel costs (if public transport) and parking cost and availability (if car). Store

1. Meat Shopping

Store Information: Address & Type of Store	Ratings: Price	Ratings: Quality	Ratings: Convenience	Usual Mode: Mode (Normal method of travel)	Usual Mode: Time (Estimated travel time for normal mode (minutes))	Alternative Modes: Modes (Modes that would be used if normal mode was unavail.)	Alternative Modes: Times (Estimated travel times for these modes (minutes))
Usual Store Coles Supermarket, Burnside	3	3	1	car dr	4	1 walk 2 3 4	8
Alternative Stores 1 Woolworths Supermarket, Burnside	3	3	1	car dr	4	1 walk 2 3 4	8
2 Turners Butchers, Burnside	4	2	2	car dr	6	1 walk 2 3 4	10

2. Major Grocery Shopping

Store Information: Address & Type of Store	Ratings: Price	Ratings: Selection	Ratings: Convenience	Usual Mode: Mode (Normal method of travel)	Usual Mode: Time (Estimated travel time for normal mode (minutes))	Alternative Modes: Modes (Modes that would be used if normal mode was unavail.)	Alternative Modes: Times (Estimated travel times for these modes (minutes))
Usual Store Coles Supermarket, Burnside	3	2	1	car dr	4	1 Walk 2 3 4	8
Alternative Stores 1 Woolworths Supermarket, Burnside	2	3	1	car dr	4	1 walk 2 3 4	8
2 Half Case Supermarket, Glenunga	4	3	3	car dr	10	1 2 3 4	

Fruit and Vegetable Shopping

Store Information: Address & Type of Store	Ratings: Price	Ratings: Quality	Ratings: Convenience	Usual Mode: Mode (Normal method of travel)	Usual Mode: Time (Estimated travel time for normal mode (minutes))	Alternative Modes: Modes (Modes that would be used if normal mode was unavail.)	Alternative Modes: Times (Estimated travel times for these modes (minutes))
Usual Store Coles Supermkt, Burnside	2	3	1	car dr	4	1 walk 2 3 4	8
Alternative Stores 1 The Fruit Bowl, Burnside	4	2	2	car dr	4	1 walk 2 3 4	8

FIGURE 4.2: EXAMPLE OF SOME MODE AND STORE INFORMATION COLLECTED FROM SHOPPING QUESTIONNAIRE

location was coded using the zonal scheme constructed for the 1977 Metropolitan Adelaide Data Base Survey (Pak Poy and Associates 1978) and then further identified using a special code for each store within a zone. This method of coding locations allowed ready utilisation of the 1977 MADBS network information.

Importantly the same coding scheme was used in the diaries as in the shopping questionnaire. This permitted the two data (sub)sets to be linked. An example diary page is shown in Figure 4.3. The final format of the diary was the result of extensive pilot testing. The dimensions of the diary (200 mm x 140 mm) were chosen to encourage the respondent to carry it and record activities as they occurred. The address and sample number of the household was inscribed inside the front cover, as well as a means of identifying the respondent (normally christian name) and the interviewer's and project leader's names and telephone numbers.

Also inside the front cover was the day the respondent was to commence recording his/her activities. This information was repeated on the first blank diary page (to be filled in by the respondent). Pages 2 and 3 contained some 'commonly asked questions' about the survey, complete with answers. These related to the aims of the survey, reasons why certain items of information were needed and confidentiality. For example, answers were supplied on why information on activities was wanted for an entire week, the usefulness of information on in home activities and how to record activities 'I regard as private'. These questions were considered crucial in alleviating doubts some respondents may have had in supplying, possibly sensitive, information. Following these questions were three pages of instructions, an example diary, and immediately before the blank diary pages, a page containing nothing except, in bold black type, three reminder points; namely, to record all travel - even minor trips, to record each shop or building visited at non-home destinations, and to carefully read the example diary.

IF YOU TRAVELLED TODAY

(Fill in for each trip)

TIME TRIP STARTED	TIME TRIP FINISHED	METHOD OF TRAVEL (Tick One): CAR DRIVER	CAR PASSENGER	CAR POOL	MOTOR CYCLE	TAXI	TRAIN	BUS	BICYCLE	WALK	TIME SPENT WALKING TO AND FROM CAR OR PUBLIC TRANSPORT (MINUTES)	IF PUBLIC TRANSPORT USED: FARE (CENTS)	NUMBER OF TRANSFERS	WAITING TIME (MINUTES)	IF CAR OR MOTORCYCLE USED: NUMBER OF PEOPLE (INCL. DRIVER)	PARKING (Tick 1): STREET PARKING	OFF STREET PARKING	HOME OFF STREET	PARKING COST (CENTS)
7.55	8.27							✓			8	40¢	-	2					
1.00	1.03									✓									
1.05	1.08									✓									
5.00	5.23		✓								4				2		✓		—
5.35	5.40		✓								1				2			✓	
7.20	7.35		✓												2			✓	
9.45	10.02		✓												2			✓	

EXAMPLE DIARY

GENERAL ACTIVITIES

TODAY IS *Thursday*

TIME ACTIVITY STARTED	TIME ACTIVITY FINISHED	WHAT WERE YOU DOING?	WAS A CHILD UNDER 12 WITH YOU AT THE TIME: YES	NO	HOW OFTEN DO YOU DO THIS (Tick One): ABOUT DAILY	ABOUT WEEKLY	IRREGULARLY	NONE OF THESE	IF DAILY OR WEEKLY ACTIVITY Do you usually do this at about this time? (Tick One): YES	NO	Varies	WHERE WERE YOU?	MONEY SPENT ON ACTIVITY: $	c
midnight	6.45	sleeping		✓	✓				✓			home		
6.45	7.05	breakfast		✓	✓				✓			"		
7.05	7.35	housework		✓	✓				✓			"		
7.35	7.55	dressing		✓	✓				✓			"		
7.55	8.27	bus to work		✓	✓				✓					
8.27	1.00	worked		✓	✓				✓			University of Adelaide North Terrace		
1.00	1.03	walked to deli		✓			✓					Snack Bar Rundle Mall		
1.03	1.05	bought lunch		✓			✓					near Pultney St.	1	56
1.05	1.08	back to office		✓			✓							
1.08	1.50	lunch		✓	✓				✓			Workplace		
1.50	5.00	worked		✓	✓				✓			"		
5.00	5.23	travelled to shops		✓			✓					Coles-Supermarket		
5.23	5.35	grocery shopping.		✓			✓					Burnside Shopping Centre	6	20
5.35	5.40	travelled home		✓										
5.40	6.10	prepared meal		✓	✓				✓					
6.10	7.05	dinner		✓	✓				✓					
7.05	7.20	cleaned up		✓	✓				✓					
7.20	7.35	travelled to friends		✓		✓					✓			
7.35	9.45	socializing		✓		✓					✓	Finniss St, North Adelaide		
9.45	10.02	travelled home		✓		✓					✓			
10.02	12.00	bed		✓	✓				✓					

FIGURE 4.3: EXAMPLE DIARY PAGE

Blank diary pages were divided into two parts. The lower half was designed to facilitate personal documentation of the nature, time and (if non-travel) place of each activity episode. In addition information on the regularity of participation for each activity, expenditure on the activity and whether a child under 12 years old was present with the respondent was requested. The upper half was designed to allow the respondent to provide further information on each trip undertaken (i.e. travel activity). In content this trip information represented a subset of data items typically included in a conventional travel survey. For all trips, method of travel was to be recorded. Mode specific information requested included direct trip costs (fare for public transport and parking cost for car travel), access and egress walk times (for public transport and car), wait time and number of transfers (for public transport) and parking type and number of occupants (for car).

Response rates for the survey are shown in Table 4.1. The original sample consisted of 534 private dwellings. Households in 49 (nine per cent) of these dwellings could not be contacted. Another 68 (i.e. 14 per cent of 485) households refused to supply any information, except whether they were resident there in 1977. Of the remaining 417 households, 179 were fully participating while 238 supplied partial information. 356 households supplied socio economic data and valid data for at least one section of the shopping questionnaire.* Information collected in the activity diaries totalled to 3,431 person days, containing 13,847 reported trips and 47,876 reported activity episodes. Sixty three per cent of contacted households had participated in the MADBS HIS, with the remaining 32 per cent moving in since 1977.

---

* The number of usable responses varied between food shopping categories, ranging from 356 for major grocery shopping to 326 for minor grocery shopping.

TABLE 4.1

HOUSEHOLD CATEGORISED RESPONSES TO 1980 ADELAIDE ACTIVITY DIARY SURVEY

Response Category	Number of Households
Fully participating households	179
Households with partial returns	238
Total refusals	68
No contact	49
Total	534

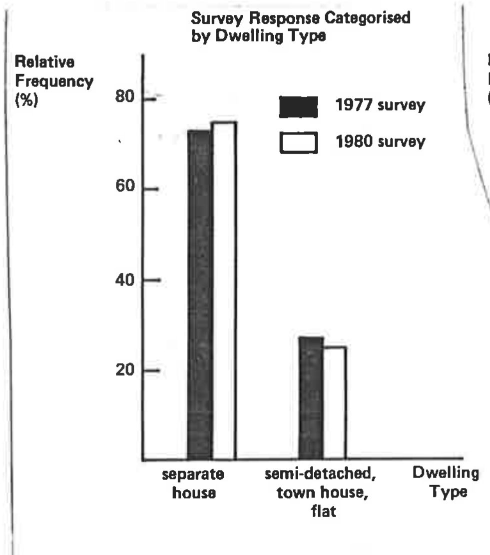
The comparatively low response rates are hardly surprising given the extremely detailed and arduous nature of the survey.* They do, however, raise the concern that the sample is not representative enough, even for the exploratory type analyses reported in this and forthcoming chapters. An extensive range of checks were conducted, measuring the representativeness of the 1980 sample against MADBS HIS information. Some of these checks are shown in Figures 4.4 and 4.5. These figures shown differences between ATDATAS respondents and a spatially equivalent set of MADBS HIS respondents with respect to household and personal characteristics. It is apparent from these figures that only minor differences exist in the socio-demographic mix of the two samples. It is unlikely that the socio-demographic differences exhibited would contribute significantly to a change in shopping behaviour.

It needs to be stressed that it was not the sole objective, or even a major objective, of ATDATAS to furnish data for the analyses presented in this study. The aims of ATDATAS were wider than this. Accordingly, the data has been used by a number of researchers in a variety of contexts. The data hs been used by the South Australian Department of Transport. Other examples of data use may be found in papers by Barnard (1985, 1986), the extensive work of Clarke et al. (1985) and minor utilisation by Wigan (1982). The present study only uses a small portion of the data collected; principally, that obtained in the shopping questionnaire and some of the shopping activities reported in the diaries.

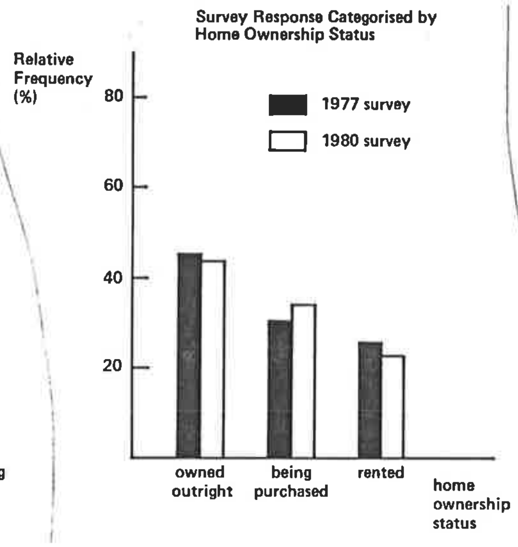
In analyses reported in this and subsequent chapters, segments of the total amount of shopping information collected were used, depending upon the particular requirements of the models developed. The subsets used are as follows:

---

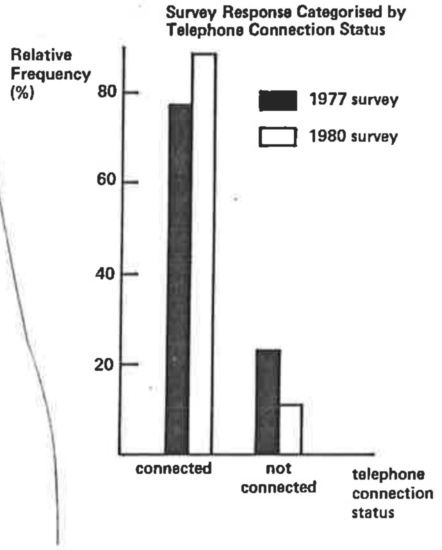
* The response rates are broadly consistent with those obtained in the Banbury activity diary survey, conducted by the Transport Studies Unit, Oxford University (Jones et al. 1983).



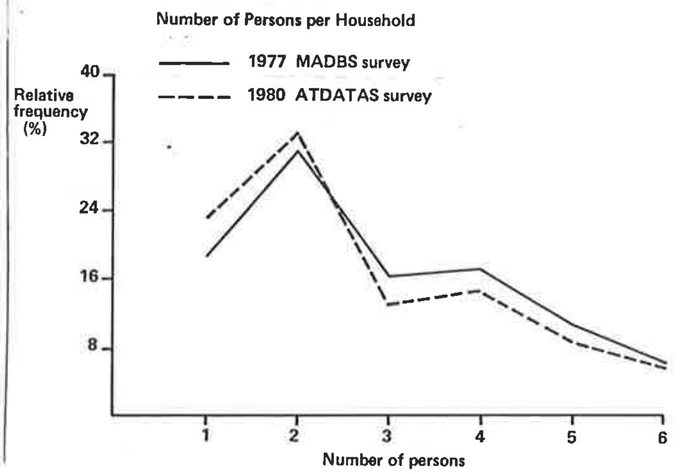
Survey Response Categorised by Dwelling Type
Relative Frequency (%)
80
60
40
20
1977 survey
1980 survey
separate house
semi-detached, town house, flat
Dwelling Type



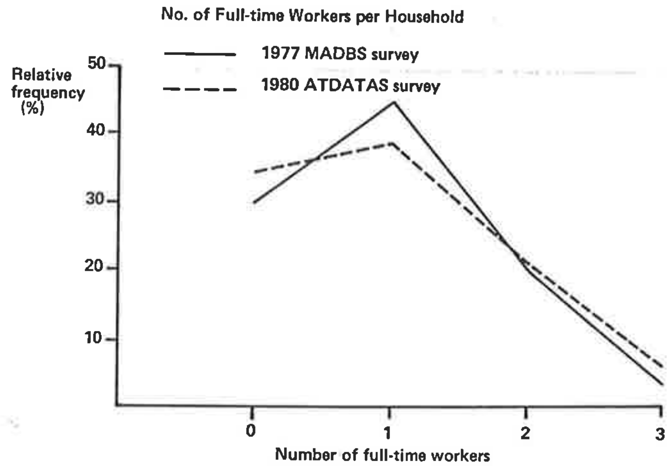
Survey Response Categorised by Home Ownership Status
Relative Frequency (%)
80
60
40
20
1977 survey
1980 survey
owned outright
being purchased
rented
home ownership status



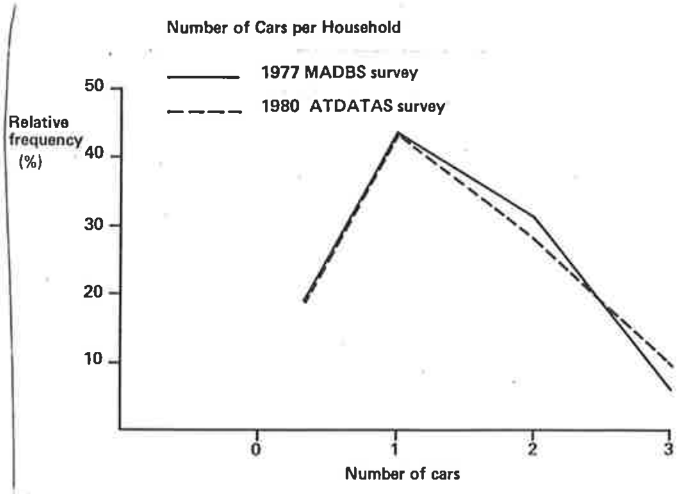
Survey Response Categorised by Telephone Connection Status
Relative Frequency (%)
80
60
40
20
1977 survey
1980 survey
connected
not connected
telephone connection status



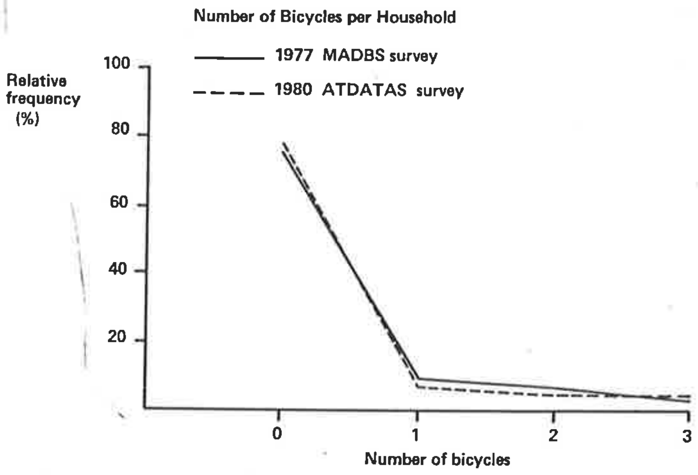
Number of Persons per Household
1977 MADBS survey
1980 ATDATAS survey
Relative frequency (%)
40
32
24
16
8
1
2
3
4
5
6
Number of persons



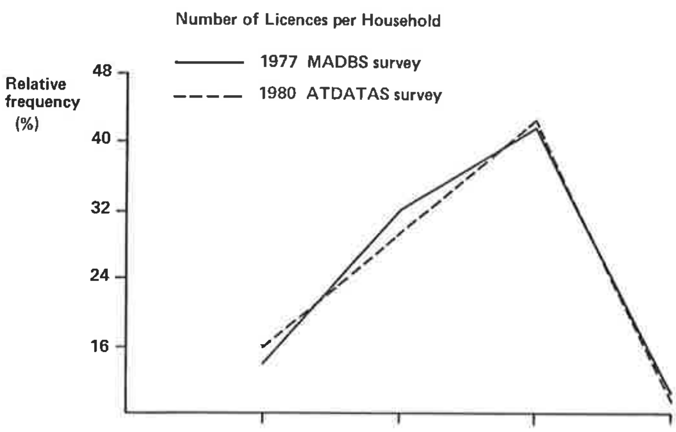
No. of Full-time Workers per Household
1977 MADBS survey
1980 ATDATAS survey
Relative frequency (%)
50
40
30
20
10
0
1
2
3
Number of full-time workers



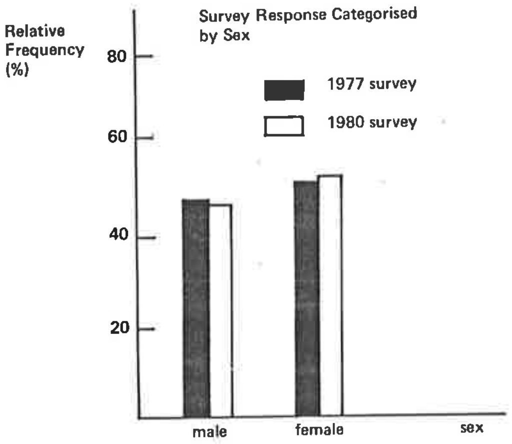
Number of Cars per Household
1977 MADBS survey
1980 ATDATAS survey
Relative frequency (%)
50
40
30
20
10
0
1
2
3
Number of cars

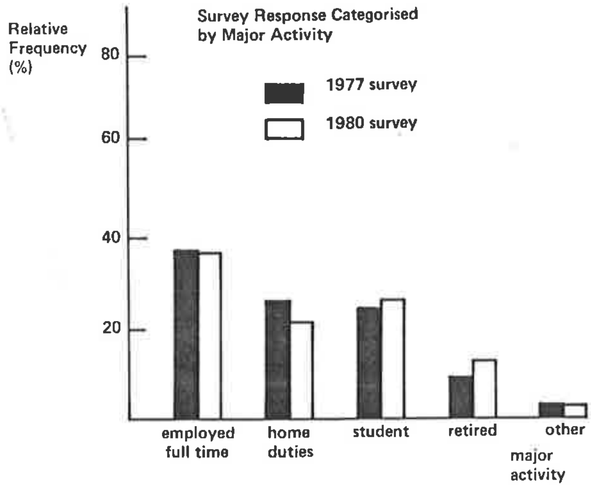


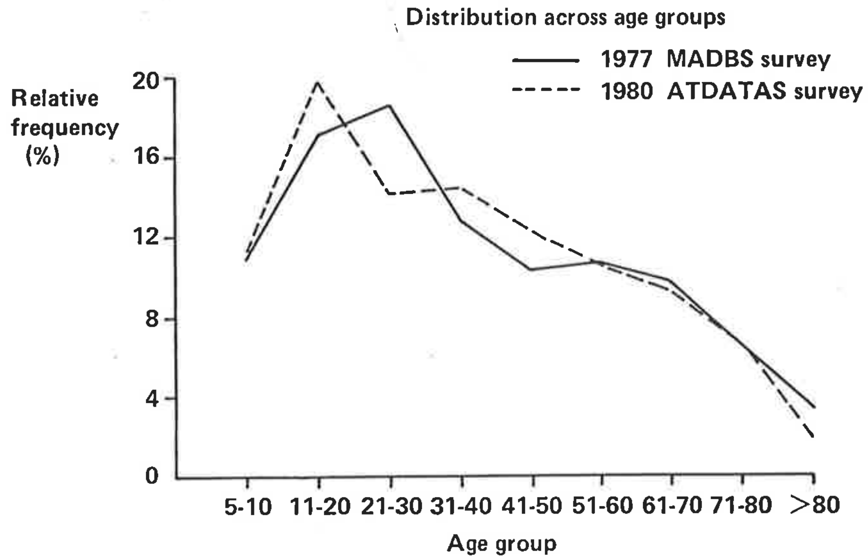
Number of Bicycles per Household
1977 MADBS survey
1980 ATDATAS survey
Relative frequency (%)
100
80
60
40
20
0
1
2
3
Number of bicycles



Number of Licences per Household
1977 MADBS survey
1980 ATDATAS survey
Relative frequency (%)
48
40
32
24
16







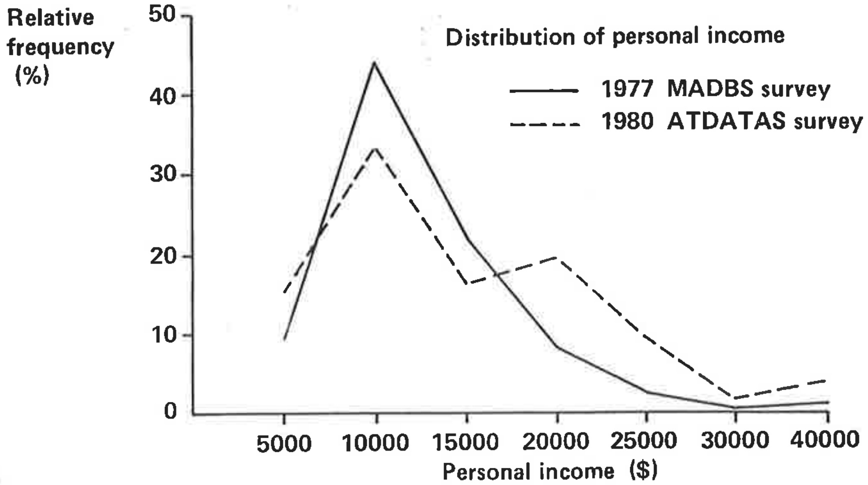


FIGURE 4.5 : ANALYSIS OF SURVEY RESPONSE BY PERSON CHARACTERISTICS

(i) In the remainder of this chapter, all respondents who supplied information for meat, major grocery, minor grocery or greengrocery shopping, respectively, in the shopping questionnaire.

(ii) In Chapter 5, which reports on choice set specification, all respondents who supplied information for major grocery shopping.

(iii) In Chapter 7, which reports on shopping travel patterns, all respondents who supplied information for meat, major grocery and greengrocery shopping with 'home' as the only connecting activity both for chosen and non-chosen stores.

(iv) In Chapter 7, which reports on estimation of integrated store/expenditure choice model, all respondents (main household shoppers) who supplied income data, information for major grocery shopping in the shopping questionnaire and who had recorded in the diaries a grocery shopping activity, with positive expenditure, to a store which formed one of the stores mentioned by the respondent in the grocery section of the shopping questionnaire.

The rationale for the restrictions imposed in obtaining each of the data subsets will become apparent in discussion pertaining to the analyses. Further information on ATDATAS procedures, etc. can be found in the survey documentation (Barnard 1981).

3. BASIC MODE/DESTINATION CHOICE MODELS FOR FOUR CATEGORIES OF FOOD SHOPPING

3.1 GENERAL DATA ANALYSIS

A general analysis of destination choice data collected is provided in Tables 4.2 to 4.4.

For major grocery shopping in all a total of 43 supermarkets were mentioned as either being used or as possible alternatives to where respondents presently shopped. However any one respondent mentioned only a maximum of six supermarkets. Frequency tabulations and an analysis of price, selection and convenience ratings for each supermarket chain are supplied in Table 4.2. As is evident from this table, statistically significant differences existed between some of the supermarket chains with Half Case rating best on the price dimension and Target supermarkets significantly better on selection. However, respondents apparently perceived considerable diversity between stores within a supermarket chain with differences in means often statistically significant (contrast with Louviere and Meyer 1981, pp. 414).* Considerable diversity also existed between respondents' perceptions, with standards deviations of store attributes expressed as a percentage of their mean, generally in the range of 20-30 per cent.

For meat and greengrocery shopping, frequency counts and information relating to destination attributes for classes of stores are displayed in Tables 4.3 and 4.4 respectively. Not surprisingly supermarkets were perceived as being more homogeneous on both price and quality dimensions than either meat or fruit shops (from a comparison of standard deviations). Although prices at

---

* Unless otherwise stated statistical significance is measured at the five per cent level.

TABLE 4.2

COMPARISON OF ATTITUDINAL DESTINATION ATTRIBUTE MEASURES FOR DIFFERENT SUPERMARKET CHAINS

Store Chain	Frequency		Price		Selection		Convenience	
	Chosen	Alternative	Mean	s.d.	Mean	s.d.	Mean	s.d.
Action Price	9	12	1.69	0.61	1.59	0.36	3.00	1.14
Big Heart	3	4	2.86	1.43	2.14	1.01	3.43	1.18
Coles	83	202	1.97	0.41	2.27	0.57	2.54	1.08
Foodland	32	76	2.01	0.53	2.21	0.66	2.60	1.29
Half Case	7	31	1.05	0.20	2.50	0.62	1.69	0.66
Serv Wel	4.	10	2.79	0.66	1.80	0.57	3.13	1.77
Target	58	63	1.81	0.60	2.87	1.32	2.60	1.24
Tom the Cheap	19	32	1.92	0.42	2.11	0.50	2.65	1.29
Woolworths	116	139	1.77	0.48	2.23	0.70	2.57	1.09
Others	25	46	2.13	0.47	1.07	0.29	1.73	0.78

TABLE 4.3

COMPARISON OF ATTITUDINAL DESTINATION ATTRIBUTE MEASURES FOR DIFFERENT MEAT STORE TYPES

MEAT SHOPPING

Store type	Frequency Chosen	Frequency Alternative		Attribute Ratings Price	Attribute Ratings Quality	Attribute Ratings Convenience
Meat shops	218	273	mean	3.07	3.89	3.67
			s.d.	(.83)	(.82)	(1.05)
Supermarkets	115	252	mean	2.99	3.21	3.60
			s.d.	(.65)	(.70)	(1.01)

Note: s.d. = standard deviation.

TABLE 4.4

COMPARISON OF ATTITUDINAL DESTINATION ATTRIBUTE MEASURES FOR DIFFERENT GREENGROCERY STORE TYPES

GREENGROCERY SHOPPING

Store type	Frequency			Attribute Ratings		
	Chosen	Alternative		Price	Quality	Convenience
Central Market	25	34	mean	1.95	4.22	2.86
			s.d.	(.63)	(.82)	(1.35)
Fruit Shops	199	171	mean	3.08	3.79	3.68
			s.d.	(.80)	(.81)	(.93)
Supermarkets	117	196	mean	3.01	3.14	3.71
			s.d.	(.63)	(.71)	(.95)

Note: s.d. = standard deviation.

butchers and fruiterers were rated as being slightly higher than at supermarkets, the hypothesis that there were no differences in perceived prices could not be rejected at the five per cent significance level ($t = 1.93$ for meat shopping and $t = 1.24$ for greengrocery shopping). Respondents did, however, perceive greengrocery prices to be lower and quality better at Adelaide's Central Market than at other outlets (price and quality differences tested at the 5% significance level). Further, meat and fruit shops were rated significantly higher than supermarkets on quality.

Modal information collected is analysed in Table 4.5. From this Table, for major shopping purposes, approximately 75 per cent of respondents used a motor vehicle (either as driver or passenger). The only other method of travel of any significance was walking (which was the usual mode for approximately 20 per cent of respondents). This ranking was preserved for alternative modes, but with minor modes such as public transport slightly increasing in relative terms (i.e. being viewed more often as an alternative mode than as the usual method of travel). For minor grocery shopping trips walking assumed greater importance, being the reported usual mode for 45 per cent of individuals surveyed. Nevertheless, even for minor grocery shopping trips, travel by car was still the usual method of travel for the majority of respondents.

Also shown in Table 4.5 are journey times for each mode. A notable feature of a previous study of grocery shopping travel choices (Kostynuik 1975) was that the mean home-to-usual-shop travel time for those walking or going by car was virtually identical (9.9 minutes and 9.8 minutes respectively). This was seen as further evidence for the constant travel time budget hypothesis (Jones 1978). Kostynuik's results, however, are not duplicated in the Adelaide data set with mean walking time being significantly greater than mean car travel time for all shopping purposes.

TABLE 4.5

MODAL INFORMATION : VARIOUS FOOD SHOPPING TRIP PURPOSES

Mode	Usual Mode %	Alternative Mode %	Travel Time: Usual Mode mean	s.d.	Travel Time: Alternative Mode mean	s.d.
Major Grocery Data Set						
car driver	221 (62)	488 (48)	6.46	4.07	7.56	4.30
car passenger	55 (15)	133 (13)	6.81	3.73	8.02	5.11
public transport	10 (3)	75 (7)	15.30	7.91	15.30	9.42
bike	3 (1)	20 (2)	8.50	2.12	9.25	4.71
walk	60 (17)	286 (28)	9.16	5.99	11.96	6.80
home delivery	8 (2)	20 (2)				
Minor Grocery Data Set						
car driver	147 (45)	211 (37)	4.02	2.31	4.60	3.32
car passenger	18 (6)	50 (9)	4.40	3.64	5.73	4.44
public transport	7 (2)	31 (5)	13.75	11.09	14.68	12.92
bike	8 (2)	37 (7)	5.83	4.88	5.21	4.16
walk	147 (45)	237 (42)	6.40	4.66	10.01	7.43
Meat Data Set						
car driver	196 (58)	288 (38)	7.60	5.77	6.94	4.48
car passenger	47 (14)	92 (12)	8.11	6.31	8.00	5.93
public transport	9 (3)	81 (11)	15.57	13.04	16.08	13.43
bike		23 (3)	-	-	9.58	5.31
walk	71 (21)	251 (33)	9.51	6.53	11.72	6.41
home delivery	16 (5)	15 (2)				
Greengrocery Data Set						
car driver	216 (60)	235 (37)	7.54	6.86	7.02	4.82
car passenger	51 (13)	85 (13)	8.83	9.34	8.71	8.30
public transport	9 (3)	60 (9)	14.22	2.66	16.38	11.40
bike	2 (1)	20 (3)	7.00	-	8.74	4.56
walk	67 (20)	226 (35)	10.06	6.52	11.93	6.74
home delivery	9 (3)	15 (2)				

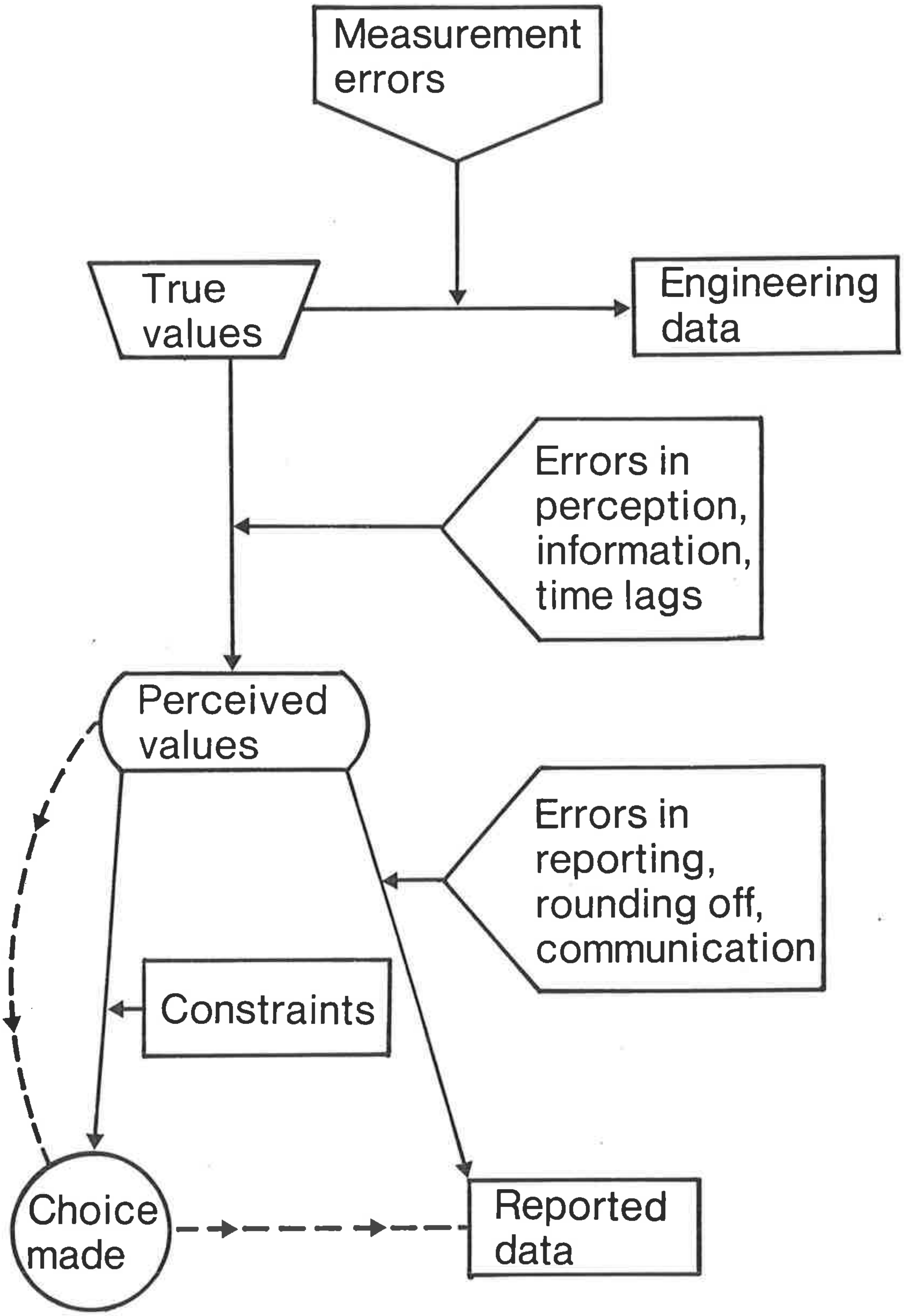
Note: s.d. = standard deviation.

3.2 RELATIONSHIP BETWEEN REPORTED AND NETWORK TRAVEL TIME VALUE

Journey times were further analysed by examining the relationship between reported and network time measurements. Network measures were available for highway times and distances from the 1976 Metropolitan Data Base Study (Pak Poy and Associates 1978). Evidence from the South Australian Highways Department suggested that highway conditions had not substantially changed between 1976 and 1980.

Figure 4.6 which is taken from Ortuzar (1982) diagramatically depicts the relationships between reported, perceived, true and measured values of travel attributes. Most research has concentrated on the relationship between perceived (reported) and true attribute values (e.g. O'Farrel and Markham 1974, Levin et al. 1979). These studies have been confined to the work trip. The technique generally used is to obtain estimates of the true value by careful manual coding of journeys or by some other means, e.g. to have interviewers with stop watches trace respondents' routes. These values are then compared to travel times reported by respondents. Although it is possible to identify a number of distorting influences on reported times, such as forcing attribute scales onto the respondent that may not correspond with his thinking about the decision, problems of rounding and post purchase bias (Daly 1978), overall this research has demonstrated a strong correspondence between true and reported values. A consistent result, however, has been that commuters have a tendency to overestimate travel times. There is also some limited evidence that attitudinal and situational factors affect time perceptions.

Substantially less research has been conducted on the relationship between the true levels of travel attributes and their engineering estimates. A study of note in this area is by Talvitie and Dehgani (1979). They



Notional relationship between choice and different attribute measurements

(Source : Ortuzar, 1982)

FIGURE 4.6

concluded that network measurements failed to meet standards for acceptable accuracy. Analysis showed low correlation coefficients between the two measures, especially for public transport component times. Models estimated on the two sets of data yielded differences in coefficient estimates that were statistically significant.

Formally the relationships outlined in the above discussion can be represented by a system of equations:

$$R_i = f_{1i}\ (H_i),\ i = 1,\ 2\ \ldots,\ L \qquad (4.1)$$

$$H_i = f_{2i}\ (T_i),\ i = 1,\ 2,\ \ldots,\ L \qquad (4.2)$$

$$N_i = f_{3i}\ (T_i),\ i = 1,\ 2,\ \ldots,\ L \qquad (4.3)$$

where R_i is the reported value for the ith attribute, H_i is the perceived value, T_i is the true value and N_i is the network value.

Provided N_i is a monotonic function of T_i, equation (4.3) can be rearranged as $T_i = f_{4i}\ (N_i)$. Then, by substitution,

$$R_i = f_{1i} \left(f_{2i} \left[f_{4i}\ (N_i) \right] \right)$$

or,

$$R_i = g_i\ (N_i),\ i = 1,\ 2,\ \ldots,\ L \qquad (4.4)$$

which is the relationship that is further explored here.

Results from regressions, constructed to analyse regularities in reported travel times are shown in Tables 4.7a and 4.7b, with the variable definitons provided in Table 4.6. The reported travel times analysed were home to store car travel times and walking times. Independent variables consisted of highway network measures, a variable indicating whether data related to a chosen or non-chosen mode, shopping travel related variables and socio-economic

TABLE 4.6

VARIABLE MNEMONICS AND DEFINITIONS FOR THE REGRESSION ANALYSES OF TABLES 4.7A and B.

MNEMONIC	VARIABLE DEFINITION
AGELT25	a binary variable taking value 1 if the respondent's age is less than 25 years and 0 otherwise
AGE2535	a binary variable taking value 1 if the respondent's age is between 25-35 years, inclusive, and 0 otherwise
AGE3660	a binary variable taking value 1 if the respondent's age is between 36-60 years, inclusive, and 0 otherwise
CAR DRIVER	a binary variable taking value 1 if travel to the shopping activity is as car driver and 0 if travel is as car passenger (constructed to test for car travel time reporting differences between car drivers and car passengers)
CHOICE	variable constructed to test for differences in travel time reporting for chosen and non-chosen alternatives, taking a value of 1 if chosen alternative and 0 if non-chosen alternative
EDUCATED	a binary variable taking value 1 if the respondent was tertiary educated and 0 otherwise.
PEAK	a binary variable taking value 1 if the shopping activity is normally done on Thursday or Friday nights or Saturday morning and 0 otherwise
SEX	a binary variable taking value 1 if the respondent is a male and 0 if the respondent is a female
SHOPPING FREQUENCY	monthly shopping frequency
TTIMEN	network travel time estimate . For car travel times TTIMEN is obtained directly from MADBS network data . For walking times the estimate is obtained from MADBS network highway distances and an assumed walking speed of 5'/sec.
WORKING	a binary variable taking value 1 if the respondent is full or part-time employed and 0 otherwise.

characteristics. Previous research (Michaels 1974) had suggested the functional relationships to be generally non-linear. With this in mind the specified form for the regression equations was;

$$R_i = \delta_0 + \delta_1 N_i^{\delta_2} + \sum_{\ell=3}^{L+2} \delta_\ell O_{\ell-2} \qquad (4.5)$$

where O_2, O_2, ..., O_L are socio economic and travel related binary variables and δ' (= δ_0, δ_1, ..., δ_{L+2}) is a vector of parameters.

The overall shape of the function specified in equation (4.5) depends crucially on the value of δ_2. For values of $\delta_2 > 1$, the function relating reported values to network values will increase at an increasing rate, and for values of δ_2 between 0 and 1 increase at a decreasing rate (see Figure 4.7). In the event for walk-ing times, as a result of convergence problems encountered in attempting to estimate the non-linear regressions, resort was made to linear regression*. Further, for the reported car travel time regressions δ_2 tended to take a value about one and the hypothesis that the linear form was correct for all shopping categories could not be rejected at the 95% significance level.**

Examining Tables 4.7A and B it can be seen that the highway network measures were always statistically significant at or above the 10% level and took on the postulated sign, indicating a positive association between reported times and their corresponding network measures. The parameter values associated with the highway network

---

* No attempt was made to trace the cause for non-convergence since these analyses represented only a very minor part of the study. Another potentially useful technique that could be used in analysing the relationship between reported and objective travel times is the method of cubic-splines (e.g. Hensher 1984a).

** Since the power functional form includes linearity as a special case (corresponding to $\delta_2 = 1$) an F-test can be applied to test for difference between models; see, for example, Kmenta 1971, pp. 446-468.

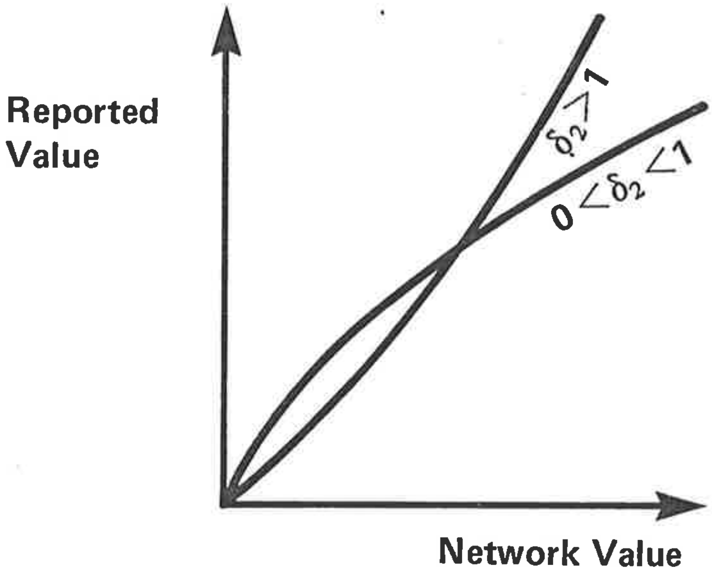


FIGURE 4.7: GRAPH OF POWER FUNCTION $R_i = N_i^{\delta_2}$ FOR DIFFERENT VALUES OF δ_2

TABLE 4.7A

NON-LINEAR REGRESSION ANALYSES OF PERCEIVED CAR TRAVEL TIME FOR VARIOUS FOOD SHOPPING PURPOSES

VARIABLE LIST	FOOD SHOPPING CATEGORIES							
	Major Grocery		Minor Grocery		Meat		Greengrocery	
	δ	s.e.	δ	s.e.	δ	s.e.	δ	s.e.
TTIMEN	1.063**	0.2408	0.5958**	0.0613	0.9686**	0.2591	0.9317**	0.3728
CHOICE	-0.3175	0.2518	-0.3270	0.3299	-0.1183	0.3183	-0.0929	0.4897
PEAK	0.3688	0.2486			0.5460*	0.3309	0.0008	0.5248
CAR DRIVER	-0.1022**	0.2871	0.4506	0.4265	0.4275	0.3700	0.6485	0.5899
SHOP FREQUENCY	0.3688**	0.1720			0.0670	0.0964	0.5248	0.4339
AGELT25	-0.7488	0.4776	-1.6414**	0.6449	-0.3838	0.6910	1.6200*	0.9656
AGE2535	-0.9114**	0.3772	-1.2012**	0.5219	-0.8124	0.5297	-0.5270	0.8038
AGE3660	-0.3163	0.3126	-0.6592	0.4204	-0.4199	0.4272	0.1724	0.6432
SEX	0.0354	0.3725	-0.2597	0.5554	0.4009	0.5449	-0.9697	0.7743
EDUCATED	0.2983	0.2802	0.1787	0.3681	-0.2719	0.3701	-0.5990	0.5800
WORKING	-0.0335	0.2514	0.1308	0.3446	0.4855	0.3510	0.0540	0.5474
POWER PARAMETER, δ_2	1.0423**	0.0865			1.0517**	0.0897	1.1092**	0.1395
CONSTANT, δ_0	2.4921**	0.6799	3.4945**	0.4372	2.5836**	0.7006	1.8785	1.2747
Number of observations	778		324		506		492	
Sum of squared residuals	7000.81		2472.91		5000.83		10361.93	

Notes:
1. Form of regression analyses is expressed in equation (4.5)
2. Variables are defined in Table 4.6
3. s.e. = standard error
4. * indicates that the parameter is significant at the 90% level and ** indicates that the parameter is significant at the 95% lev
5. Linear regression analysis only for minor grocery shopping
6. Dependent variable = reported car travel time.

TABLE 4.7B

LINEAR REGRESSION ANALYSES OF PERCEIVED WALKAING TIMES FOR VARIOUS FOOD SHOPPING PURPOSES

VARIABLE LIST	FOOD SHOPPING CATEGORIES							
	Major Grocery		Minor Grocery		Meat		Greengrocery	
	δ	s.e.	δ	s.e.	δ	s.e.	δ	s.e.
TTIMEN	0.0691*	0.0372	0.1156**	0.0406	0.2018**	0.0399	0.1527**	0.0472
CHOICE	-3.3968**	1.0716	-3.9057**	0.7420	-1.7626*	1.0241	-2.0467*	1.0623
SHOP FREQUENCY	0.9587*	0.5255			0.2690	0.2508	1.9859**	0.6698
AGELT25	-2.9721**	1.4249	-0.9498	1.4099	-2.6200	1.7092	-2.0481	1.6160
AGE2535	-1.9277**	1.2543	0.4683	1.1908	-1.4673	1.3223	-2.0940	1.4367
AGE3660	-0.6550	1.0146	-0.1474	0.9556	-1.6128*	0.9950	-1.7384	1.1099
SEX	-0.7520	1.1132	-0.5220	1.1015	-1.9918	1.2442	-2.0155	1.3524
EDUCATED	-0.3226	0.8887	-0.1485	0.8451	-1.3180	0.9417	-0.1670	0.9568
WORKING	-2.4871**	0.8632	-1.6912**	0.8092	-0.1165	0.8959	-1.7883*	0.9591
CONSTANT	11.1721**	1.7009	9.4270**	1.0603	8.2647**	1.3681	7.9506**	1.8142
R^2	0.121		0.118		0.158		0.132	

Notes:
1 Variables defined in Table 4.6.
2. s.e. - standard error.
3. * indicates the parameter is significant at the 90% level and ** indicates that parameter is significant at the 95% level.
4. Dependent variable = reported walking time.

times were always about 1 indicating no systematic divergence between reported car travel times and their corresponding network measures. Conversely, parameters attached to the network measures related to walking times were always much less than 1. It is apparent from this that there exists a large divergence between the network measure constructed to estimate walking times and perceived times. Either respondents are considerably underestimating true times or the network measure considerably overestimating these times.

'CHOICE' was a dummy variable constructed to test a possible tendency for individuals to over report travel times for non-chosen alternatives in comparison to chosen alternatives (Daly 1978, Stopher and Meyburg 1975). The negative sign of the 'CHOICE' parameter in all regressions is indeed indicative of support for such an hypothesis; however, for car times it is insignificantly different from zero. Furthermore for walking times, given the generally poor fit of these regressions, statistical significance of this parameter may merely demonstrate utility maximising behaviour (i.e. choosing the alternative with the lowest travel time) rather than providing conclusive evidence for over-enthusiasm on the part of respondents in attempting to justify current modal use.

'PEAK' is another dummy variable taking a value of 1 if shopping was usually done at peak times (i.e. Thursday or Friday nights or Saturday mornings) and 0 if on weekdays. The, as postulated, positive sign on its parameter estimate implies that network measures do not capture important time related variations. For two shopping purposes, major grocery and meat, this coefficient is significant at the 10 per cent level using a one tailed t-test, which is perhaps surprising given the variable's crude construction.

Parameter estimates associated with the other shopping travel related variables, 'MODE' (1 for car driver, 0 for car passenger) and 'FREQUENCY', were generally statistically insignificant for the car regressions. However, for some shopping walking trips the coefficients of the shopping frequency variable were significantly negative, signalling a tendency by those who shop less often to inflate travel times.

Statistically significant amongst the socio-economic variables were some age dummies and working status. The size and sign of parameter estimates attached to the age dummies should be interpreted relative to the base group (age > 60). That these parameters generally took on a negative sign therefore suggests that younger people tend to report lower travel times than the elderly. This result, of course, may have a basis in real walking speed differences between the age groups. Similarly, the negative sign on some worker dummy variables indicates that those who work may underestimate some walking times relative to those not working. The statistical significance of these socio-economic variables casts doubt on the unqualified use of objective time measures in predicting travel behaviour. In particular this adds to evidence that the relationship between perceived and actual travel times may systematically vary with socio-economic characteristics. As these characteristics change over time, so will travel time perceptions and hence travel behaviour. Any model not accounting for this influence will therefore give inaccurate forecasts.

In summary, the correspondence between network and reported measures for car trips was acceptable rather than good, and for walking trips was poor. Other research outlined suggests, in these circumstances, reported measures will be better at explaining travel behaviour and in any case provide the closest approximation to the true values. Therefore for most of the models estimated in this study reported values were used. In general, network values were only used when reported values were unavailable as, for instance, in the shopping pattern model of Chapter 6 and the choice set analysis of Chapter 5.

3.3 ANALYSIS OF ATTITUDINAL DESTINATION ATTRACTIVENESS MEASURES

It was also decided to used perceived destination attributes in preference to objective measures such as floor space or employment. The case for using perceived measures has been argued in Chapter 2 and need not be repeated. An interesting feature that emerged in the analysis of these attributes concerned the variances associated with the perceived destination attribute measures. In particular, ex-ante expectations were that selection of goods would be more accurately perceived than store prices and therefore would display less variance. Data analysis revealed no significant differences in variance between these variables.* A possible explanation is that respondents tended to give average type answers for those attributes about which they were more uncertain. It is also to be noted that no statistically significant differences emerged when comparing the price, selection/quality attribute ratings for chosen and non-chosen alternatives (see Table 4.8).

3.4 VARIABLE DEFINITIONS

Table 4.9 gives the variables, their codes and definitions that are used in the models presented throughout the remainder of this report.

It can be seen from this table, that the quality variables, identified by the B_i vectors in Chapter 3, are represented by perceived selection or quality of goods

---

* The test involved application of the F-distribution. F-values were calculated by:

$$F = \frac{n_x s_x^2 / (n_x - 1)}{n_y s_y^2 / (n_y - 1)}$$

where n_k is the size of the kth sample and s_k is the standard deviation of this sample.

TABLE 4.8

MEAN AND STANDARD DEVIATIONS OF DESTINATION ATTRACTIVENESS MEASURES FOR CHOSEN AND NON-CHOSEN STORES*

Meat Shopping

	Frequency	Mean Price (s.d.)	Mean Quality (s.d.)
chosen	336	2.89 (0.78)	3.92 (0.83)
non-chosen	536	3.06 (0.74)	3.40 (0.79)

Major Grocery Shopping

	Frequency	Mean Price (s.d.)	Mean Selection (s.d.)
chosen	356	2.86 (0.64)	3.49 (0.76)
non-chosen	659	3.49 (0.76)	3.21 (0.73)

Minor Grocery Shopping

	Frequency	Mean Price (s.d.)	Mean Selection (s.d.)
chosen	326	3.54 (0.85)	3.15 (0.85)
non-chosen	360	3.50 (0.85)	3.11 (0.85)

Greengrocery Shopping

	Frequency	Mean Price (s.d.)	Mean Quality (s.d.)
chosen	345	2.93 (0.83)	3.01 (0.77)
non-chosen	429	3.01 (0.86)	3.33 (0.80)

* Note: The price, quality and selection attributes were measured using the rating scale outlined in Section 2 (see also Table 4.9).

TABLE 4.9

VARIABLE MNEMONICS AND DEFINITIONS

MNEMONIC	CHAPTER(S)	VARIABLE DEFINITION
AL(1)	5	binary variable taking value 1 for alternative 1 and 0 otherwise
$AVGCONV_d$	5	averaged store convenience rating for grocery shopping at store d, $AVGCONV_d = \sum_q^{Q_d} GCONV_d$ where Q_d is the number of individuals rating store d in terms of store convenience
$AVGPRICE_d$	5	averaged price rating for grocery shopping at store d (refer to $AVGCONV_d$)
$AVGSEL_d$	5	averaged selection rating for store i (refer to $AVGCONV_d$)
CARSAV	4,6	an alternative specific index for car competition within a household taking values: CARSAV = NCARS/(NLICEMPL + 1) for the car alternative and O otherwise.
FOREIGN	4	a binary variable taking value 1 for the public transport alternative if of non-English country of birth and O otherwise
GACCESS	5	an index of accessibility to grocery stores: $GACCESS = \sum_d GDISTANCE_d$
$GCONV_d$	4,6,7	perceived convenience rating for grocery shopping at store d with the rating scale as: 1. - store convenience much below average 2. - store convenience slightly below average 3. - store convenience about average 4. - store convenience slightly above average 5. - store convenience much above average

$GDISTANCE_d$	5	distance from home to grocery store d (kms)
$GEXPEND_{qdm}$	7	observed expenditure on groceries by individual q at store d when using mode m.
$GPRICE_d$	4,6,7	perceived rating of grocery prices at store d (rating scale used similar to that for GCONV)
$GSEL_d$	4,6,7	perceived selection of grocery goods available at store d (rating scale used similar to that for GCONV)
GSPENDTRIP	5	average grocery shopping expenditure per major grocery shopping trip ($)
HINCOME	7	average hourly household income per employed household member (cents)
INCLVAL	4,5	inclusive value term
LICENCE	4	a binary variable taking value 1 for car as an alternative if a driver's licence is held and 0 otherwise
LIFECYCLEYC	6	binary life cycle variable taking value 1 for shopping pattern alternatives with a tour arrangement linking all stores together if children less than five years old are present in the household and 0 otherwise.
M	4,5	binary variable taking value 1 for shopping alternatives involving home delivery of goods and 0 otherwise
M1	4,5	binary variable taking value 1 for those alternatives involving use of car as driver and 0 otherwise.
M2	4,5	binary variable taking value 1 for those alternatives involving use of car as passenger and 0 otherwise
M7	4,5	binary variable taking value 1 of public transport and 0 otherwise
M8	4	binary variable taking value 1 for those alternatives involving use of bicycle and 0 otherwise

M9	4,5	binary variable taking value 1 for those alternatives involving walking and 0 otherwise
$MCONV_d$	4,6	perceived convenience rating for meat shopping at store d (rating scale used similar to that for GCONV)
$MPRICE_d$	4,6	perceived rating of grocery prices at store d (rating scale used similar to that for GCONV)
$MQUALITY_d$	4,6	perceived quality of meat available from store d (rating scale used similar to that for GCONV)
$NCARS_q$	5	number of cars available in the household pertaining to individual q
NLICEMPL	4,6	number of driver's licences held by employed household members
$NGSTORES_q$	5	number of stores in the reported grocery shopping choice set of individual q
$NUMBERLOCS_{sp}$	6	number of different shopping centres involved in shopping pattern sp
$NUMBERTOURS_{sp}$	6	number of tours involved in shopping pattern sp
$OCONV_d$	4	perceived convenience rating for miscellaneous food shopping at store d (rating scale used similar to that for GCONV)
$OPRICE_d$	4	perceived rating of miscellaneous food prices at store d (rating scale used similar to that for GCONV)
$OSEL_d$	4	perceived rating of miscellaneous food prices at store d (rating scale used similar to that for GCONV)
RESYEARS	5	years of household residency at the surveyed address
SCLEE	7	selectivity correction factor calculated from the Heckman/Lee method
SECEDUC	5	binary variable taking value 1 if no more

		than secondary education was achieved and 0 otherwise
$TCOST_i$	4,7	travel cost associated with alternative i: if bus involved TCOST = reported bus fare, if car involved TCOST = network highway distance x 0.12, if walk or bike only involved TCOST = 0 (cents)
$TTIME_i$	4,7	perceived home - destination travel time associated with alternative i (minutes)
$TCOSTN_i$	5,6	travel cost associated with alternative i as calculated solely from network values (cents)
$TTIMEN_i$	5,6	travel time associated with alternative i as calculated from network values (minutes)
$VCONV_d$	4,6	perceived convenience rating for fruit and vegetable shopping at store d (rating scale used similar to that for GCONV)
$VPRICE_d$	4,6	perceived rating of fruit and vegetable prices at store d (rating scale used similar to that for GCONV)
$VQUALITY_d$	4,6	perceived quality of fruit and vegetables available from store d (rating scale used similar to that for GCONV)
WORKER	4	binary variable taking value 1 for the car driver alternative if working status is full or part time and 0 otherwise

NOTES: 1. To assist explanation subscripts are on occasions used in this table. They have generally been dropped, however, when presenting results.

2. The store attribute ratings are referred to generally throughout the text as PRICE, SEL, CONV. The perceived price rating is sometimes referred to by p* (see notation).

within a store and perceived store convenience. Thus,

$$\psi_{iq} = \exp\ (\gamma_1\ SEL_{iq} + \gamma_2\ CONV_{iq}) \qquad (4.6)$$

Store prices are also measured as perceived by the respondent rather than representing actual prices. To circumvent this data deficiency, a transformation between real and perceived prices is assumed:

$$p_i = p_i\ (p^*_{iq}) = a_1\ p^{*\ a_2}_{iq} \qquad (4.7)$$

where p^*_{iq} is the perceived price level at store i by individual q, and a_1 and a_2 are unknown parameters. No additive constant was included in this transformation on the basis that zero priced goods are perceived to be free. * Note that although the perceived price, selection/quality and convenience variables are strictly measured on an interval scale, past evidence (e.g. Louviere et al. 1979, Louviere and Meyer 1981) suggests that they may be successfully used as though ratio scaled.

Other variables included in Table 4.9 are travel-related variables, socio-economic descriptors and mode-specific dummy variables. The socio-economic variables are a subset of those included in model development, with variables excluded which proved to be insignificant for all models.

3.5 ESTIMATION RESULTS FOR BASIC SIMULTANEOUS MODE/DESTINATION FOOD SHOPPING CHOICE MODELS

The first modelling step in the study was to estimate some basic simultaneous mode/destination choice models for the four categories of food shopping. In line with virtually all past research, the conditional indirect utility

---

* The difference between perceived and actual prices really only becomes important in the work reported in Chapter 7.

expressions of equation (3.20) were approximated by a linear in the parameters function:

$$V_{iq} = \alpha_1 a_2 \log p^*_{iq} + \alpha_1 \gamma_1 SEL_{iq} + \alpha_1 \gamma_2 CONV_{iq}$$

$$+ \alpha_2 (T - t_{iq}) + \alpha_3 (HINCOME_q - c_{iq})$$

$$+ \sum_{\ell=4}^{L+3} \alpha_\ell O_{\ell-3} + \alpha_1 \log a_1 + \epsilon_{iq} \qquad (4.8)$$

where the vector $(O_1, O_2, \ldots, O_L)$ contains additional socio-economic and mode-specific dummy variables. The ϵ_{iq} were assumed to be distributed iid extreme value type 1 so the choice model was of the multinomial logit form:

$$\text{Prob}\left\{ I_q = j \right\} = \exp(\overline{V}_{jq}) \;/\; \exp(\overline{V}_{qi}) \qquad (4.9)$$

where the $\overline{V}_{qi}$ are the representative portion of conditional indirect utility functions and the alternatives for individual q ($i = 1, 2, \ldots, N_q$) defined jointly by modes and destinations.

The modal alternatives examined were limited to car driver, car passenger, bus, bicycle and walk. Taxi trips were omitted because they were so few in number. Rail travel is not an available alternative in the eastern and north-eastern suburbs of Adelaide. Destination alternatives were included as individual stores. Estimation was thus achieved using elemental alternatives rather than some spatial aggregation of these (e.g. number of destination types in a traffic zone).

Explanatory variables fall into two main categories, generic variables and alternative specific variables. All variables must be defined with respect to the dependent variable (i.e. have some testable relationship with choice (Hensher 1979)). Generic variables take on different values for each alternative, whereas alternative specific variables have values of zero for some alternatives. From Table 4.9 generic level of service variables included in the models are travel time (TTIME) and trip cost (TCOST). Generic destination attractiveness variables used were price, selection, quality and convenience. These were also combined with certain socio-economic attributes, but these combinatory variables proved statistically insignificant in explaining choices. Other socio-economic variables were included in alternative specific form, being added onto the conditional indirect utility expressions for one or more modes.

The significance of individual variables may be assessed by reference to their t-statistics. Two measures are provided for overall model goodness of fit. One, the likelihood ratio index, is defined as:

$$\rho^2 = 1 - L_c^*/L_o^*$$

where L_c^* is the log likelihood for the fitted model and L_o^* is the log likelihood for the 'at-equal-shares' hypothesis. As more of the data variance is explained by the model, the log likelihood at convergence becomes a smaller negative number and ρ^2 increases. The likelihood ratio index therefore behaves similarly to the correlation coefficient used in regression. A difference is that values of ρ^2 of 0.2 to 0.4 represent excellent fit, while such values for R^2 would tend to represent indifferent to poor fit (McFadden 1979). The other measure is 'percentage correctly predicted'. This measure simply scores a '1' when the chosen alternative is also that alternative with the highest predicted probability and a '0' otherwise. Scores are then summed and expressed as a percentage of all observations.

Results are displayed in Tables 4.10 - 4.13. As can be seen from these tables, all variables took on their postulated sign. For the destination attractiveness descriptors, there was found to be a negative relationship

TABLE 4.10

MAJOR GROCERY SHOPPING MODE/STORE CHOICE:
BASIC SIMULTANEOUS LOGIT MODEL RESULTS

Variable Name*	Parameter estimate	Standard error	T - Statistic
log (GPRICE)	-1.851	0.3934	-4.71
GSEL	0.9219	0.1444	6.39
GCONV	0.9494	0.1332	7.13
TTIME	-0.0575	0.0122	-2.65
TCOST	-0.0178	0.0085	-1.51
M	-0.8561	0.9632	-0.89
M2	-0.6548	0.7741	-0.85
M7	-1.295	0.9097	-1.42
M8	-0.6847	1.080	-0.63
M9	-0.7885	0.7678	-1.03

ρ^2	0.275
% correctly predicted	
- at zero	29
- at convergence	59

* Note: variables defined in Table 4.9.

TABLE 4.11

MINOR GROCERY SHOPPING MODE/STORE CHOICE
BASIC SIMULTANEOUS LOGIT MODEL RESULTS

Variable Name*	Parameter estimate	Standard error	T - Statistic
log (OPRICE)	-0.5772	0.4808	-1.20
OSEL	0.1368	0.1463	0.93
OCONV	0.7297	0.1739	4.20
TTIME	-0.0512	0.0244	-2.10
TCOST	-0.0492	0.0165	-2.98
M2	-1.334	0.4143	-3.22
M7	-0.3561	0.8136	-0.44
M8	-1.257	0.5219	-2.41
M9	-0.6743	0.2402	-2.81

ρ^2	0.140
% correctly predicted	
- at zero	37
- at convergence	54

* Note: Variables defined in Table 4.9.

TABLE 4.12

MEAT SHOPPING MODE/STORE CHOICE: BASIC SIMULTANEOUS LOGIT MODEL RESULTS

Variable Name*	Parameter estimate	Standard error	T - Statistic
log (MPRICE)	-1.799	0.4131	-4.35
MQUALITY	0.9960	0.1430	6.97
MCONV	0.4870	0.1240	3.93
TTIME	-0.0184	0.0153	-1.20
TCOST	-0.0131	0.0055	-2.38
CARSAV	1.688	0.5335	3.17
WORKER	1.089	0.4207	2.59
M	1.645	0.7453	2.21
M2	0.5463	0.5585	0.98
M7	-0.4722	0.6697	-0.70
M9	0.4831	0.5217	0.93
ρ^2	0.261		
% correctly predicted			
- at zero	33		
- at convergence	61		

* Note: Variables defined in Table 4.9.

TABLE 4.13

GREENGROCERY SHOPPING MODE/STORE CHOICE
BASIC SIMULTANEOUS LOGIT MODEL RESULTS

Variable Name*	Parameter estimate	Standard error	T - Statistic
log (VPRICE)	-0.9553	0.3720	-2.57
VQUALITY	0.6454	0.1222	5.28
VCONV	0.7019	0.1371	5.12
TTIME	-0.0116	0.0118	-0.98
TCOST	-0.0210	0.0086	-2.42
CARSAV	2.107	0.7659	2.75
WORKER	1.388	0.5377	2.58
FOREIGN	3.201	1.266	2.53
M	0.0881	1.054	0.08
M2	0.6358	0.7415	0.86
M7	-1.565	1.291	-1.21
M8	-0.5125	1.199	-0.43
M9	0.4115	0.7196	0.57

ρ^2	0.265
% correctly predicted	
- at zero	36
- at convergence	62

* Note: Variables defined in Table 4.9.

between perceived high prices and choice of destination. Conversely as ratings of selection and convenience increased, the probability of selecting the destination also increased. An interesting feature is that price and selection ratings are statistically insignificant for minor grocery shopping trips. It appears for these trips, for which shopping expenditure is minimal, individuals consider only convenience and travel related aspects (providing an empirical basis for the term 'convenience shopping'). Major grocery shopping trips are also dominated by convenience considerations. However, for meat shopping, quality of meat conditions selection of store more than store prices or convenience. Similarly, quality of fruit and vegetables ranks above convenience in importance when selecting a greengrocery store.

Travel cost proved to be significant, in statistical terms, in explaining choices in all but one of the food shopping categories. Travel time was only statistically significant in the two grocery shopping models. For a linear specification of the conditional indirect utility functions of equation (3.9), the income and total time variables drop out when comparing mode/destination combinations i and j.

To investigate the cause for the marginal significance of travel time in some of the models, this variable was divided into in-vehicle and out-of-vehicle times, constructed from network data. In-vehicle time consisted of car time and line haul public transport time. The main component of out-of-vehicle travel time was walking time. In general out-of-vehicle time proved highly significant in statistical terms, but in-vehicle time was in every case statistically insignificant and for some models even took the incorrect sign. A possibility was that in-vehicle time was interacting with store convenience. This was not substantiated, however, by an examination of correlation coefficients and exclusion of the convenience variable did not greatly increase the t-value associated with in-vehicle travel time. Since a positive sign associated with in-vehicle time is theoretically implausible it was decided to adopt the more constrained form for inclusion of travel time and drop the separate terms measuring in-vehicle and out-of-vehicle time.

A generally pleasing element of the models is the statistical insignificance of most modal dummies. This implies that no important characteristics intrinsic to a given mode have been omitted. (A similar conclusion can be drawn from the statistical insignificance of store type related dummies that were included in development of the final model forms.) Evidence provided by Talvitie and Kirshner (1978) is that for many models modal specific variables contribute 60 per cent or more of explained utility. Hensher (1981) has argued that use of such models in a forecasting context is suspect because of problems with new modes and possible changes in excluded characteristics of existing modes.

The principal socio-economic effects are reflected in the car availability and working status variables. The variable measuring car availability was designed to capture competition for cars within a household. It is defined as:

$$CARSAV = NCARS/(NLICEMPL + 1)^{*}$$

where NCARS is the number of cars available to the household and NLICEMPL is the number of employed household members with a licence. Other socio-economic variables such as sex, age and licence status were found to be statistically insignificant.

As noted, the dependent variable used in the above models was defined jointly by modes and destinations. This

---

* The number '1' was added to the denominator to prevent dividing by '0' and to give weight to those households with one or more cars but no members who were both licensed and employed.

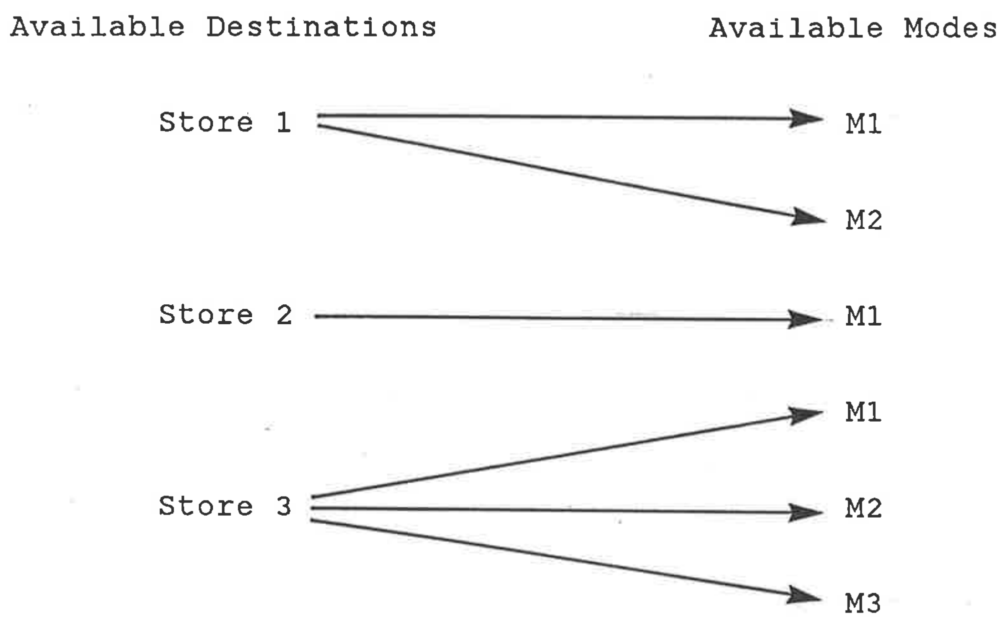


FIGURE 4.8 : ILLUSTRATION OF THE CHOICE SET FOR A HYPOTHETICAL INDIVIDUAL

conforms to simultaneous estimation of mode and destination choices. As reviewed in Chapter 2, a more general structure involves nesting mode choice within destination choice. The nested logit model allows for a pattern of dependence between the unobserved attributes of alternatives and as a consequence does not suffer from some of the restrictions inherent in the 11A axiom (which is a property of the simple multinomial logit model used in simultaneous estimation).

Notwithstanding this theoretical superiority, a loss of efficiency does result in the standard two-staged nested estimation procedure due to estimates of estimates associated with the inclusive value terms being passed on from one model to another. With small data sets, such as the one used in the current study, this problem may become severe. Also it is unclear how the procedure should be applied with variable choice set data. Figure 4.8 illustrates a hypothetical choice set that might be observed in the data under study. Say the individual depicted in Figure 4.8 chooses store 1 and mode 1 and that modes 1, 2 and 3 represent the global modal choice set. Then a nested logit model may be estimated and inclusive value terms calculated as:

$$IV_d = \log \sum_{m}^{3} \exp\,(\alpha_1 t_{dm} + \alpha_2 c_{dm})$$

where, again, a generic formulation for the time and cost variables has been used and d = 1, 2, 3. The question arises about how to calculate inclusive value terms for destinations 1 and 2 which exhibit some modal unavailability for the individual depicted. If very high values of t_{dm} and c_{dm} are used (in an endeavour to reflect extremely high disutility associated with the unavailable modes and hence, extremely low probabilities of choice) then the distribution of IV_d may take on an odd trimodal shape. Also just one mode being unavailable would make IV_d so small compared to the situation with all modes available, it may be difficult for the model to distinguish between degrees of modal unavailability. Another interesting observation is that if the modal availability situation between stores 1 and 2 had been reversed, the individual would have been excluded from the lower (mode choice) level of the nest. To the author's knowledge these issues have received no attention in the literature*. For these reasons in this study the simultaneous mode/destination structure was favoured. There is limited evidence to suggest that the use of a simple multinomial logit model is not likely to result in substantial errors when compared with more general structures (Williams and Ortuzar 1979).

For completeness, however, a nested logit model was estimated for major grocery shopping. Results from the two-staged estimation are shown in Table 4.14. Again all variables take on their anticipated signs. The coefficient of the inclusive value term being positive but less than 1, lies within a range consistent with utility maximising theory (McFadden 1978). That it takes a value close to 1 (in fact, indistinguishable from 1 at normal confidence levels) suggests that the simultaneous model structure mainly utilised in this thesis, is supported by the data.*

3.6 POLICY SIMULATION

The models have been used to simulate modal and destination changes resulting from possible policy actions (using the sample enumeration method) and these are shown in Tables 4.15 and 4.16.

---

* In a forthcoming article Hensher (1986) examines some of the difficulties that variable choice sets impose on the sequential estimation of nested logit models. This article also compares results from sequential nested logit (S-NL) estimation with full information maximum likelihood estimation.

* The 95% confidence interval for the inclusive value coefficient of Table 4.14 is 0.5462 - 1.0742.

TABLE 4.14

MAJOR GROCERY SHOPPING MODE/STORE CHOICE: NESTED LOGIT MODEL RESULTS

Mode Choice Model

Variable Name	Parameter Estimate	Standard Error	T-Statistic
TTIME	-0.0950	0.0395	-2.41
TCOST	-0.0094	0.145	-0.66
M	-1.026	0.8466	-1.21
M2	-2.501	0.4556	-0.55
M7	-3.598	0.7659	-4.70
M8	-2.571	0.8971	-2.87
M9	-1.992	0.4337	-4.59

ρ^2	0.505
% correctly predicted	
- at zero	48
- at convergence	84

Store Choice Model

Variable Name	Parameter Estimate	Standard Error	T-Statistic
log (GPRICE)	-1.867	0.4065	-4.59
GSEL	0.9826	0.1489	6.60
GCONV	0.8820	0.1301	6.78
INCLVAL	0.8102	0.1347	6.01

ρ^2	0.335
% correctly predicted	
- at zero	37
- at convergence	71

TABLE 4.15

MODEL PREDICTIONS OF CHANGES IN MODAL USE
RESULTING FROM SIMULATED POLICY CHANGES

Policy Action	Predicted Percentage Change in Mode Use			
	car driver	car passenger	bus	walk
1. Major Grocery Shopping				
10% decrease in bus fares	-0.4	-0.8	+2.5	-1.3
10% increase in car costs	-0.5	-1.0	+0.5	+2.1
10% decrease in bus travel times	-0.4	-0.8	+3.1	-1.3
10% increase in car travel times	-0.6	-1.3	+0.7	-2.6
2. Minor Grocery Shopping				
10% decrease in bus fares	0.0	0.0	2.0	-0.1
10% increase in car costs	-0.4	-0.7	+0.1	+0.6
10% decrease in bus travel times	-0.1	0.0	+2.0	-0.1
10% increase in car travel times	-0.4	-0.8	+0.5	+0.6
3. Meat Shopping				
10% decrease in bus fares	-0.1	-0.2	+2.5	-0.4
10% increase in car costs	-0.1	-0.4	+0.5	+1.4
10% decrease in bus travel times	-0.1	-0.4	+4.0	-3.3
10% increase in car travel times	-0.2	-0.8	+1.0	+1.0
4. Greengrocery Shopping				
10% decrease in bus fares	0.0	-0.1	+2.3	-0.2
10% increase in car costs	-0.1	-0.3	+0.7	+0.4
10% decrease in bus travel times	-0.3	-0.7	+0.7	+1.3

TABLE 4.16

AVERAGE DIRECT STORE CHOICE ELASTICITIES

Variable	Elasticity Estimate			
	major grocery shopping	minor grocery shopping	meat shopping	greengrocery shopping
price	-0.611	-0.300	-0.617	-0.306
quality/selection	1.038	0.182	1.304	0.765
convenience	1.197	1.276	0.627	0.839

It is apparent from Table 4.15 that mode use for food shopping trips is predicted to be extremely unresponsive to any conventional policy change. Elasticities are extremely low, even lower than those usually found in commuter mode choice studies. What little mode switching that may occur as a result of policy initiatives is predicted to be primarily between walk and car or walk and public transport (possibly combined with a change of destinations) rather than between public transport and car. This pattern of cross elasticities in part reflects perceived and physical constraints binding on travellers as revealed in their stated choice sets. Brog and Erl (1981) in a study of possible changes in West German public transport patronage from reducing public transport times and/or increasing car costs, using a substantially different methodological approach (but one which emphasised the role of constraints), also predicted the effects on walk and bicycle trips to be as great if not greater, than the effect on public transport or car trips. The oft recurrence of this pattern has also been noted in a review of British and European 'before and after' studies of policy changes designed to encourage increased public transport patronage (Jones et al. 1980).

Direct store choice elasticities for the perceived price, quality/selection and convenience variables are shown in Table 4.16. These are average elasticities; that is, direct elasticities averaged across all stores in the data sets. In the main these too are less than one, although there are exceptions, notably, elasticities relating to quality of store merchandise for meat shopping, store convenience for minor grocery shopping and selection of items for major grocery shopping.

4. SUMMARY

This chapter has provided an introduction to the empirical setting of this study. The basic models reported in this chapter provided a reasonably good account of food shopping behaviour within a broad economic theoretic framework. The models, however, may be further refined. In forthcoming chapters three areas of refinement are examined. In the next chapter, the first of these, pertaining to choice set specification, is considered.

APPENDIX 4A: THE ATDATAS SHOPPING QUESTIONNAIRE

FORM 3: SHOPPING INFORMATION

SAMPLE NUMBER ____________ PERSON NUMBER ____________

MEAT SHOPPING

1. Do you do most of the meat shopping or does another member of the household?
   - (i) R main meat shopper ☐
   - (ii) another member of the household main meat shopper (enter person number) Go to Q. 7 ☐☐
   - (iii) other (specify) ____________

2. Usually about how often do you buy meat for the household?
   - (i) 2–3 times per week ☐
   - (ii) about once a week ☐
   - (iii) about once a fortnight ☐
   - (iv) about once a month ☐
   - (v) other (specify) ____________

3. A. When do you usually shop for meat?
   - (i) weekdays ☐
   - (ii) Thursday night ☐
   - (iii) Friday night ☐
   - (iv) Saturday morning ☐

   B. If for some reason you could not shop on [usual time] at what other times would you possibly go?
   - (i) weekdays ☐
   - (ii) Thursday night ☐
   - (iii) Friday night ☐
   - (iv) Saturday morning ☐

4. Normally about how much does the household spend on meat each week?
   $ ☐☐

5. A. Is meat ever delivered to the home?
   - (i) most meat home delivered ☐
   - (ii) occasionally meat home delivered ☐
   - (iii) no meat home delivered—Go to Q.6 ☐

   IF R ANSWERS (i) or (ii) to Q. 5A

   B. How often on average is meat delivered to the home?
   - (i) about once a week ☐
   - (ii) about once a fortnight ☐
   - (iii) about once a month ☐
   - (iv) other (specify) ____________

   C. Do you usually order the meat over the telephone or do you go to the shops to do this?
   - (i) telephone order ☐
   - (ii) shop order ☐
   - (iii) mostly telephone but sometimes go to the shops and order there ☐
   - (iv) mostly order at shops but sometimes telephone ☐
   - (v) other (specify) ____________

   D. Names and addresses of firms who deliver meat (if possible)

   Firm 1 ____________

   ____

   Firm 2 ____________

   ____

   E. Normally how much is the delivery charge (cents) ☐☐☐

6.

ADDRESS AND TYPE OF SHOP	RATINGS			JOURNEY TYPE	USUAL MODE INFORMATION			ALTERNATIVE MODES			ONLY CAR	
	Price	Quality	Convenience	How R normally fits travelling there into his/her pattern (special trip from home, after work, etc.)	Mode: Normal method of travel	Time: Estimated travel time for normal mode (minutes)	Fare: If normal mode taxi or public transport —estimated fare (c)	Modes: Modes that would be used if normal mode was unavail.	Times: Estimated travel times for these modes (minutes)	Fares: If alt. modes taxi or public transport —estimated fares (c)	Park Cost	Park Avail
Usual shop								1				
								2				
								3				
								4				
Alternative shops 1.								1				
								2				
								3				
								4				
2.								1				
								2				
								3				
3.								1				
								2				
								3				

GROCERY SHOPPING

7. Do you do most of the grocery shopping or does another member of the household?
   - (i) R. main grocery shopper ☐
   - (ii) another household member—main grocery shopper (enter person number). Go to Q. 13 ☐☐
   - (iii) other (specify) ______

8. A. When you buy groceries for the household do you usually buy—
   - (i) just what you need every couple of days or so ☐
   - (ii) or have a major shopping trip every week or fortnight or so and on occasions small items purchased in between ☐
   - (iii) other (specify) ______

   B. [If (ii)]—Usually about how often do you have major grocery shopping trips?
   - (i) every week ☐
   - (ii) every fortnight ☐
   - (iii) every three weeks ☐
   - (iv) every month ☐
   - (v) other (specify) ______

9. A. When do you usually shop for groceries?
   - (i) weekdays ☐
   - (ii) Thursday night ☐
   - (iii) Friday night ☐
   - (iv) Saturday morning ☐

   B. If for some reason you could not shop on [usual time] at what other times would you possibly go?
   - (i) weekdays ☐
   - (ii) Thursday night ☐
   - (iii) Friday night ☐
   - (iv) Saturday morning ☐

10. Normally about how much does the household spend on groceries each week? $ ☐☐

11. A. Are groceries ever delivered to the home?
   - (i) most groceries home delivered ☐
   - (ii) occasionally groceries home delivered ☐
   - (iii) no groceries home delivered ☐ Go to Q. 12

   IF ANSWERS (i) or (ii) to Q. 11.A.

   B. How often on average are groceries delivered to the home?
   - (i) about once a week ☐
   - (ii) about once a fortnight ☐
   - (iii) about once a month ☐
   - (iv) other (specify) ______

   C. Do you usually order the groceries over the telephone or do you go to the shops to do this?
   - (i) telephone order ☐
   - (ii) shop order ☐
   - (iii) mostly telephone but sometimes go to the shops and order there ☐
   - (iv) mostly order at shops but sometimes telephone ☐
   - (v) other (specify) ______

   D. Names and addresses of firms who deliver groceries (if possible)

   Firm 1 ______

   ______

   Firm 2 ______

   ______

   E. Normally how much is the delivery charge (cents) ☐☐☐

12.

ADDRESS AND TYPE OF SHOP	RATINGS			JOURNEY TYPE	USUAL MODE INFORMATION			ALTERNATIVE MODES			ONLY CAR	
	Price	Selection	Convenience		Mode	Time	Fare	Modes	Times	Fares	Park Cost	Park Avail.
				How R normally fits travelling there into his/her pattern (special trip from home, after work, etc.)	Normal method of travel	Estimated travel time for normal mode (minutes)	If normal mode taxi or public transport —estimated fare (c)	Modes that would be used if normal mode was unavail.	Estimated travel times for these modes (minutes)	If alt. modes taxi or public transport —estimated fares (c)		
Usual grocery shop								1				
								2				
								3				
								4				
Alternative shops 1.								1				
								2				
								3				
2.								1				
								2				
								3				
3.								1				
								2				

ODDS AND ENDS GROCERY SHOPPING

13. A. Do you have milk delivered to the home?
   - (i) Yes ☐
   - (ii) No—Go to Q. 14 ☐

IF YES

B. On average how many times each week is milk home delivered?
   number ☐

C. Usually about how many bottles/cartons would you get each week?
   number ☐☐

14. A. Do you have bread delivered to the home?
   - (i) Yes ☐
   - (ii) No—Go to Q. 15 ☐

B. (If Yes) On average how many times each week is bread delivered to the home?
   number ☐

15. On average how often would you shop for odd items of food and groceries?
   - (i) 4—5 times per week ☐
   - (ii) 2 or 3 times a week ☐
   - (iii) about once a week ☐
   - (iv) about once a fortnight ☐
   - (v) about once a month ☐
   - (vi) less often—Go to Q. 17 ☐

16. Ask R specifically about the following (i) shops that he/she has used or would consider using for this type of shopping also used for major grocery shopping trips (ii) delicatessens the R has used or would consider using (iii) [if worker] shops where R may stop off on the way home from work (iv) any other.

ADDRESS AND TYPE OF SHOP	RATINGS			JOURNEY TYPE	USUAL MODE INFORMATION			ALTERNATIVE MODES			ONLY CAR	
	Price	Selection	Convenience	How R normally fits travelling there into his/her pattern (special trip from home, after work, etc.)	Mode: Normal method of travel	Time: Estimated travel time for normal mode (minutes)	Fare: If normal mode taxi or public transport—estimated fare (c)	Modes: Modes that would be used if normal mode was unavail.	Times: Estimated travel times for these modes (minutes)	Fares: If alt. modes taxi or public transport—estimated fares (c)	Park Cost	Park Avail
Usual shop								1				
								2				
								3				
								4				
Alternative shops 1.								1				
								2				
								3				
2.								1				
								2				
								3				
3.								1				
								2				
								3				
4.								1				
								2				
								3				
5.								1				
								2				
								3				
6.								1				
								2				
								3				

FRUIT AND VEGETABLE SHOPPING

17. Do you do most of the f. & v. shopping or does another member of the household?
    (i) R main f. & v. shopper ☐
    (ii) another member of the household main f. & v. shopper (enter person number) Go to Q.23 ☐☐
    (iii) other (specify) ______

18. Usually about how often do you buy f. & v. for the household?
    (i) 2–3 times per week ☐
    (ii) about once a week ☐
    (iii) about once a fortnight ☐
    (iv) about once a month ☐
    (v) other (specify) ______

19. A. When do you usually shop for f. & v.?
    (i) weekdays ☐
    (ii) Thursday night ☐
    (iii) Friday night ☐
    (iv) Saturday morning ☐

    B. If for some reason you could not shop on [usual time] at what other times would you possibly go?
    (i) weekdays ☐
    (ii) Thursday night ☐
    (iii) Friday night ☐
    (iv) Saturday morning ☐

20. Normally about how much does the household spend on f. & v. each week?
    $ ☐☐

21. A. Are f. & v. ever delivered to the home?
    (i) most f. & v. home delivered ☐
    (ii) occasionally f. & v. home delivered ☐
    (iii) no f. & v. home delivered—Go to Q.22 ☐

    IF R ANSWERS (i) or (ii) to Q.21A.

    B. How often on average are f. & v. delivered to the home?
    (i) about once a week ☐
    (ii) about once a fortnight ☐
    (iii) about once a month ☐
    (iv) other (specify) ______

    C. Do you usually order f. & v. over the telephone or do you go to the shops to do this?
    (i) telephone order ☐
    (ii) shop order ☐
    (iii) mostly telephone but sometimes go to the shops and order there ☐
    (iv) mostly order at shops but sometimes telephone ☐
    (v) other (specify) ______

    D. Names and addresses of firms who deliver f. & v. (if possible)
    Firm 1 ______
    ______
    Firm 2 ______
    ______

    E. Normally how much is the delivery charge (cents) ☐☐☐

22.

ADDRESS AND TYPE OF SHOP	RATINGS			JOURNEY TYPE	USUAL MODE INFORMATION			ALTERNATIVE MODES			ONLY CAR	
	Price	Selection	Convenience	How R normally fits travelling there into his/her pattern (special trip from home, after work, etc.)	Mode: Normal method of travel	Time: Estimated travel time for normal mode (minutes)	Fare: If normal mode taxi or public transport—estimated fare (c)	Modes: Modes that would be used if normal mode was unavail.	Times: Estimated travel times for these modes (minutes)	Fares: If alt. modes taxi or public transport—estimated fare (c)	Park Cost	Park Avail.
Usual fruit & veg. shop								1				
								2				
								3				
								4				
Alternative shops 1.								1				
								2				
								3				
								4				
2.								1				
								2				
								3				
3.								1				
								2				
								3				
4.								1				
								2				
								3				

23. You have said that your normal shopping pattern is to do your meat shopping at [usual meat shops], your grocery shopping at [usual grocery shops] and your greengrocery shopping at [usual fruit and vegetable shop]
   - (i) Yes ☐
   - (ii) No ☐

24. Location of usual shops (tick as appropriate)
   - (i) all at same location ☐
   - (ii) meat and grocery shops at same location, greengrocery shop at a different location ☐
   - (iii) meat and greengrocery shops at same location, grocery shop at a different location ☐
   - (iv) grocery and greengrocery shops at same location, meat shop at a different location ☐
   - (v) all at different locations ☐
   - (vi) other (specify) ____________________

   ____

25. Do you normally do all your major meat, grocery and greengrocery shopping on the same day or on different days?
   - (i) all done on same day ☐
   - (ii) meat and grocery shopping on same day, greengrocery shopping on another day ☐
   - (iii) meat and greengrocery shopping on same day, grocery shopping on another day ☐
   - (iv) grocery and greengrocery shopping on same day, meat shopping on another day ☐
   - (v) all on different days ☐
   - (vi) other (specify) ____________________

   ____

26. [If those shopping activities done on the same day are done at different locations]. Do you usually go from doing your (shopping activity) to (other shopping activities), or do you usually go home or do something else in between?
   - (i) go directly ☐
   - (ii) go home in between ☐
   - (iii) other (specify) ____________________

   ____

CHAPTER 5

CHOICE SET SPECIFICATION IN DISCRETE SHOPPING DESTINATION CHOICE MODELS: AN EMPIRICAL EXPLORATION WITH THE MAJOR GROCERY SHOPPING SUBSET.

1. INTRODUCTION

In Chapters 2 and 3 when a sample of Q individuals was considered, care was taken to subscript the set of destinations N, from which any individual's choice was made, by q. This was in recognition of the probability that different individuals choose shopping destinations from different destination choice sets. For a model to be behaviourally based, embedded in the constructs of random utility theory, the choice set used by the analyst in the estimation process should correspond to (or represent a subset of) the set of alternatives actually considered by the individual when making a choice.

In 'travel demand' applications of discrete choice models, however, choice sets have been typically specified using ad hoc procedures. For example, in mode choice it has been common practice to include transit as an alternative provided the individual resides within a certain radius of a bus line or train station, and to define auto availability on the basis of household motor vehicle ownership and possession of a driver's licence (e.g. Project Bureau for Integrated Traffic and Transport Studies 1977, Charles River Associates 1976, Richards and Ben Akiva 1975). Similarly in formulating the destination choice set (for any individual) it has been common to attribute as destination alternatives, the chosen destinations of other individuals living in the same zone (see previous references). Many studies, though, have used more loosely defined destination choice sets (e.g. Pak Poy and Associates 1978). Also interesting variations on this general criterion are apparent. For example, Adler and Ben-Akiva (1976) in their study of shopping destination choice included additional destinations 'based on deductive notions of the perception of alternatives'. In this chapter choice sets defined using ad hoc procedures, without reference to the real sets of options from which individual choices are made, are termed 'analyst assigned' choice sets.

It is the purpose of this chapter to utilize the perceived choice set data for major grocery shopping collected in the ATDATAS survey to investigate the choice set specification issue. The body of this chapter is divided into two substantive sections. In Section 2.1 it will be shown theoretically that if the choice set is not accurately defined, biased parameter estimates may result from the multinomial logit (MNL) model. This is followed in Section 2.2 by examples of different parameter estimates to emerge from mode and destination MNL models of shopping choice when reported and analyst assigned choice sets, respectively, are used in estimation. In Section 3.1 the shape and size of shopping destination choice sets are hypothesized to be a result of a search process. Empirical results are presented in Section 3.2 in support of this hypothesis. These presentations are followed by a short conclusion (Section 4).

2. EFFECTS OF CHOICE SET MIS-SPECIFICATION ON PARAMETER ESTIMATES OF MNL MODELS

This section is divided into two parts. Firstly, the effect of choice set mis-specification on MNL choice model parameter estimates is demonstrated theoretically. Secondly, some examples of different parameter estimates to result from use of reported and synthetically constructed choice sets are provided using the ATDATAS data for major grocery shopping.

2.1 EFFECT OF CHOICE SET MIS-SPECIFICATION ON MNL MODEL PARAMETER ESTIMATES: THEORY

It is possible to identify two forms of choice set mis-specification:

(a) When one or more alternatives truly considered by the individual are omitted from the choice set used in estimating the model.

(b) When one or more alternatives not considered by the individual are included in the choice set used in estimating the model.

The first type of choice set mis-specification will not affect parameter estimates when the individual choice model is of the MNL form (McFadden, 1981). This follows simply from the irrelevant alternatives (IIA) property implicit in the MNL model: that the relative odds of choosing one alternative over another is unaffected by the presence or absence of further alternatives. Provided the 11A assumption is not violated, therefore, it is perfectly proper to estimate an MNL model on a reduced choice set to that actually used by the individual.

A problem, however, does exist when the choice set used in estimation contains alternatives not considered by the individual (i.e. mis-specification exist of type b). To demonstrate the problem consider a linear-in-the-parameters conditional indirect utility function such as used in Chapter 4, $\bar{V}_{iq} = \sum_{\ell=1} \alpha_\ell \, z_{iq\ell}$. For simplicity suppose that the sample population can be segmented into two subsets, A and B. Members of subset A have all N alternatives available, whereas members of subset B do not have alternatives H + 1, H + 2, ..., N available.

Using an MNL model, the true coefficients of the conditional indirect utility functions can be estimated by maximizing the log likelihood function (Barnard and Neil 1983):

$$\Gamma = \sum_{q \in A} \sum_{i=1}^{N} k_{iq} \log P_{iq} + \sum_{q \in B} \sum_{i=1}^{H} k_{iq} \log P_{iq} \qquad (5.1)$$

where $k_{iq} = \begin{cases} 1 \text{ if } q \text{ chose } i \\ 0 \text{ otherwise} \end{cases}$

and

$$P_{iq} = \begin{cases} \dfrac{\exp(\overline{V}_{iq})}{\sum_{j=1}^{N} \exp(\overline{V}_{jq})}, & q \in A \\[2ex] \dfrac{\exp(\overline{V}_{iq})}{\sum_{j=1}^{H} \exp(\overline{V}_{jq})}, & q \in B \end{cases}$$

Suppose Γ is maximized at α^*, then:

$$\frac{\partial \Gamma(\alpha^*)}{\partial \alpha_\ell} = \sum_{q \in A} \sum_{i=1}^{N} (k_{iq} - P_{iq}) z_{iq\ell}$$

$$+ \sum_{q \in B} \sum_{i=1}^{H} (k_{iq} - P_{iq}) z_{iq\ell} = 0 \qquad (5.2)$$

Unwittingly, however, the analyst, oblivious to the unavailability of alternatives H + 1, H + 2, ..., N to members of subset B, will maximize:

$$\overline{\Gamma} = \sum_{q \in A} \sum_{i=1}^{N} k_{iq} \log \overline{P}_{iq} + \sum_{q \in B} \sum_{i=1}^{N} k_{iq} \log \overline{P}_{iq} \qquad (5.3)$$

where

$$\overline{P}_{iq} = \begin{cases} P_{iq} & , q \in A \\[2ex] \dfrac{\exp(\overline{V}_{iq})}{\sum_{j=1}^{N} \exp(\overline{V}_{jq})} & , q \in B \end{cases}$$

The relation between $\overline{P}_{iq}$ and P_{iq} for members of subset B is given by:

$$\overline{P}_{iq} = P_{iq} \left(1 - \sum_{j=H+1}^{N} \overline{P}_{jq}\right)^{(*)} \qquad (5.4)$$

Hence,

$$\frac{\partial \overline{\Gamma}(\alpha^*)}{\partial \alpha_\ell} = \sum_{q\in A} \sum_{i=1}^{N} (k_{iq} - \overline{P}_{iq}) z_{iq\ell}$$

$$+ \sum_{q\in B} \sum_{i=1}^{N} (k_{iq} - \overline{P}_{iq}) z_{iq\ell}$$

$$= \sum_{q\in A} \sum_{i=1}^{N} (k_{iq} - P_{iq}) z_{iq\ell}$$

$$+ \sum_{q\in B} \left\{ \sum_{i=1}^{H} \left[k_{iq} - P_{iq} \left(1 - \sum_{j=H+1}^{N} \overline{P}_{jq}\right)\right] z_{iq\ell}\right.$$

$$\left. + \sum_{i=H+1}^{N} \left[k_{iq} - P_{iq} \left(1 - \sum_{j=H+1}^{N} \overline{P}_{jq}\right)\right] z_{iq\ell}\right\}$$

$$= \frac{\partial \Gamma(\alpha^*)}{\partial \alpha_k} + \sum_{q\in B} \left\{ \sum_{i=1}^{H} \left[P_{iq} \sum_{j=H+1}^{N} \overline{P}_{jq}\right] z_{iq\ell}\right.$$

$$\left. + \sum_{i=H+1}^{N} \left[k_{iq} - P_{iq} \left(1 - \sum_{j=H+1}^{N} \overline{P}_{jq}\right)\right] z_{iq\ell}\right\} \qquad (5.5)$$

---

* Proof:

Since $P_{iq} = \dfrac{\exp(\overline{V}_{iq})}{\sum_{j=1}^{H} \exp(\overline{V}_{jq})}$

and $\overline{P}_{iq} = \dfrac{\exp(\overline{V}_{iq})}{\sum_{j=1}^{N} \exp(\overline{V}_{jq})}$

then,

$$\overline{P}_{iq} = P_{iq} \sum_{k=1}^{H} \frac{\exp(\overline{V}_{kq})}{\sum_{j=1}^{N} \exp(\overline{V}_{jq})} = P_{iq} \sum_{j=1}^{H} \overline{P}_{jq}$$

$$= P_{iq} \left(1 - \sum_{j=H+1}^{N} \overline{P}_{jq}\right)$$

Since,

$$\sum_{q\in B}\left\{\sum_{i=1}^{H}\left[P_{iq}\sum_{j=H+1}^{N}\overline{P}_{jq}\right]z_{iq\ell}+\left[\sum_{i=H+1}^{N}k_{iq}\right.\right.$$

$$\left.\left.-P_{iq}\left(1-\sum_{j=H+1}^{N}\overline{P}_{jq}\right)\right]z_{iq\ell}\right\}$$

will in general not be equal to zero, the parameter estimates obtained from maximizing the unrestricted choice set log likelihood function $\overline{\Gamma}$ will not coincide with those from the true log likelihood function Γ. This proof can easily be extended to the situation in which each individual has a personalized choice set.

2.2 EFFECT OF CHOICE SET MIS-SPECIFICATION ON MNL MODEL PARAMETER ESTIMATES: EMPIRICAL EVIDENCE

Having theoretically established in the previous subsection that choice set mis-specification may lead to biased parameter estimates in the MNL model, the degree and significance of such bias remains to be demonstrated in practice.

Naturally any general conclusion on this issue can only be obtained through appropriate interrogation of many data sets. Resources did not permit this to be achieved in the current study and in any case data have rarely been collected on reported choice sets. Such data, however, was collected in ATDATAS. In the remainder of this chapter, reported major grocery shopping choice sets are analyzed. In this subsection destination models estimated using reported and constructed choice sets are compared.

The models compared again utilize the linear conditional indirect utility specification of equation (4.8). They differ, however, in three aspects to the models developed in the previous chapter. Firstly, network information is used in constructing travel times and costs. Reported travel time and cost information was available only for mode/destination alternatives stated by the respondent as habiting his/her choice set. Therefore, notwithstanding the arguments forwarded in Chapter 4, it was necessary, to obtain valid comparisons, to use network level of service data in both the constructed and reported choice set models. Similarly, the perceived destination attractiveness measures were only available for stated alternatives. The second important difference between the models developed here with those presented in the previous chapter is that for each shopping destination averaged perceived price, selection/quality and convenience ratings were used rather than the individual specific values. Thirdly, and of only minor importance, sampled individuals having bicycle as a usual or alternative mode or who reported home delivery as an option for major grocery shopping were excluded from the analysis. This was because of the relatively small numbers of individuals falling into these categories (see Table 4.5). As in Chapter 4, both nested and simultaneous mode/destination structures are developed. The nested structure serves to focus attention more acutely on variations in destination choice sets.

Before comparing the models it should be recognized that reported choice sets may differ from the 'true perceived' choice sets. Provided, however, the reported choice set is a subset of the true choice set used by each individual, and the MNL specification is correct, no bias, but some inefficiency, will be introduced into the parameter estimates. The validity of the comparisons does not rely on the ability to obtain complete information on perceived choice sets.

Tables 5.1 - 5.4 contain information on nested and simultaneous mode/destination choice models estimated using respondent reported and analyst assigned choice sets. Models based on analyst assigned choice sets assumed all modes were available to all destinations for all individuals. The conclusions to be drawn from unreported models which used 'physically available' modal choice sets did not differ, however, in any substantial degree to those presented below. Analyst assigned destination choice models

TABLE 5.1

MAJOR GROCERY SHOPPING MODE/STORE CHOICE NESTED LOGIT MODEL RESULTS USING ANALYST ASSIGNED CHOICE SETS

Mode Choice Model

Variable Name*	Parameter Estimate	Standard Error	T-Statistic
TCOSTN	-0.0209	0.0191	-1.097
TTIMEN	-0.0371	0.0110	-3.405
M2	-1.516	0.2070	-7.324
M7	-2.048	0.5461	-3.750
M9	-0.4924	0.3185	-1.546
ρ^2	0.317		
% correctly predicted			
- at zero	25		
- at convergence	65		

Store Choice Model

Variable Name*	Parameter Estimate	Standard Error	T-Statistic
log(AVGPRICE)	-1.151	0.5210	-2.210
AVGSEL	1.590	0.2453	6.482
AVGCONV	-0.3322	0.1471	-2.258
INCLVAL	4.512	0.2645	17.06
ρ^2	0.470		
% correctly predicted			
- at zero	11		
- at convergence	57		

*Note: Variables defined in Table 4.9.

TABLE 5.2

MAJOR GROCERY SHOPPING MODE/STORE CHOICE
NESTED LOGIT MODEL RESULTS USING
RESPONDENT REPORTED CHOICE SETS

Mode Choice Model

Variable Name*	Estimate	Parameter Error	Standard T-Statistic
TCOSTN	-0.0110	0.0385	-0.287
TTIMEN	-0.0262	0.0204	-1.286
M2	-2.719	0.5423	-5.014
M7	-3.641	1.092	-3.335
M9	-2.288	0.6310	-3.625
ρ^2	0.539		
% correctly predicted			
- at zero	49		
- at convergence	86		

Store Choice Model

Variable Name*	Parameter Estimate	Standard Error	T-Statistic
log(AVGPRICE)	-0.0989	0.4053	-0.244
AVGSEL	0.5580	0.2167	2.575
AVGCON	0.3832	0.2252	1.701
INCLVA	1.031	0.2588	3.984
ρ^2	0.069		
% correctly predicted			
- at zero	39		
- at convergence	59		

*Note: Variables defined in Table 4.9.

TABLE 5.3

MAJOR GROCERY SHOPPING MODE/STORE CHOICE SIMULTANEOUS LOGIT MODEL RESULTS USING ANALYST ASSIGNED CHOICE SETS

Variable Name*	Parameter Estimate	Standard Error	T-Statistic
log(AVGPRICE)	-1.050	0.6009	-1.747
AVGSEL	1.566	0.2675	5.855
AVGCONV	-0.3937	0.1643	-2.396
TCOSTN	-0.1250	0.0109	-11.43
TTIMEN	-0.0786	0.0097	-8.122
M2	-1.051	0.5303	-1.982
M7	0.3542	0.4726	0.749
M9	-0.6727	0.3978	-1.691

ρ^2	0.595
% correctly predicted	
- at zero	11
- at convergence	69

TABLE 5.4

MAJOR GROCERY SHOPPING MODE/STORE CHOICE SIMULTANEOUS LOGIT MODEL RESULTS USING RESPONDENT REPORTED CHOICE SETS

Variable Name*	Parameter Estimate	Standard Error	T-Statistic
log(AVGPRICE)	-0.6302	0.6933	-0.909
AVGSEL	0.5770	0.2299	2.510
AVGCONV	0.3259	0.2355	1.384
TCOSTN	-0.0263	0.0111	-2.369
TTIMEN	-0.0107	0.0129	-0.827
M2	-2.120	0.4222	-5.022
M7	-2.932	0.6694	-4.381
M9	-1.989	0.4847	-4.105

ρ^2	0.143
% correctly predicted	
- at zero	32
- at convergence	48

*Note: Variables defined in Table 4.9.

contained the chosen store of each individual and 8 randomly selected stores from the study area.

Several points are worth noting from these tables:

(i) Superficial examination of these tables suggests that quite different models emerge from the analyst assigned and respondent reported choice sets data. Statistically many of the coefficient estimates developed on the analyst assigned choice sets data significantly differ from those obtained using respondent reported choice sets. Application of the likelihood ratio test affirms that in every case the models developed using analyst assigned choice sets are significantly different from their counterparts developed using respondent reported choice sets.

(ii) The only incorrectly valued coefficient estimates are associated with the analyst assigned choice sets data. The incorrect sign attached to 'AVGCONV' in Tables 5.1 and 5.3 suggests the presence of potentially higher utility yielding destination alternatives of which the shopper is unaware and which as a result do not form an element in the shoppers' reported choice set. Furthermore, the estimate of the 'INCLVAL' parameter (see Table 5.1), being significantly greater than 1.0, is inconsistent with random utility maximization (see McFadden 1981).*

(iii) The models developed on analyst assigned choice sets tended to be better specified, in a purely statistical sense, than those estimated using respondent reported choice sets. Typically parameter estimate standard errors were lower and overall goodness of fit measures

---

* This is a global condition. Boersch-Supan (1985) has shown as a local sufficiency, with certain restrictions being met, the parameter of inclusive value can exceed unity and be consistent with random utility maximisation.

considerably better for the analyst assigned choice sets models compared to the respondent reported choice sets models. The statistical superiority of the analyst assigned choice sets models can to some extent be attributed to the greater number of observations used in these models. Although the same set of individuals was used throughout, 9 destinations habited each analyst assigned choice set, compared to an average of 2.8 destinations in each reported choice set. Statistical superiority can also be attributed to greater variability, particularly in the travel variables, in the analyst assigned choice sets data.

The results summarized above, coupled with the theoretical exposition of subsection 2.1 suggest that a closer examination of the composition of store choice sets is warranted. In particular, although it cannot universally be concluded from the work reported above, that parameter estimates from models using analyst assigned choice sets will be significantly biased, it nevertheless appears that choice set composition exerts a significant influence on model estimates. Prior to empirically examining variations in reported major grocery shopping choice sets, a relatively simple theory is presented of store choice set determination.

3. AN INVESTIGATION OF STORE CHOICE SET DETERMINATION

It is the purpose of this section to present a theory of store choice set determination and to empirically test this theory using the grocery shopping data set for Adelaide. The theory is based on the premise that the major constraint restricting store choice sets is the stock of knowledge held by the individual regarding the urban area, which in turn reflects the amount of search he has undertaken. (This contrasts perhaps with a mode choice situation where physical constraints such as car availability may dominate.)

The importance of knowledge as a constraining influence on choice is highlighted in studies by Sheluga et al. (1979) and Devine and Marion (1979). Both these studies using substantially different methods, point to the same conclusion: that lack of knowledge may substantially limit the level of maximum utility attainable. Sheluga et al. using an experimental design found that respondents considered only a small portion of information available to them and that the chosen alternative was 'not necessarily ... the best available alternative when all information and options were taken into account' (p. 174). Of even more relevance to the subject matter here is Devine and Marion's study. Their study concerned grocery shopping behaviour and knowledge of prices. By publishing in daily newspapers information about prices on offer in various grocery stores these authors were effectively able to relax the knowledge constraint. In a post test survey 43 per cent of respondents in Ottawa-Hull, Canada, reported they had changed stores whilst the figure for Winnipeg respondents was 18 per cent. Thus a substantial number of respondents in both cities had higher utility options lying outside their choice sets. It is by including these options in the estimated choice set that biased parameter estimates result.

These arguments indicate that in choosing a store individuals follow a two step process. First they decide what store options should be searched and considered (i.e. they decide on the size and compilation of their store choice set). Only after this has been completed do they select a particular alternative. Both phases involve utility maximisation. However, the important point to note is that total utility maximisation is not unconstrained (as has been implicitly assumed). Rather information acquisition is a costly process.

Following, in subsection 3.1, economic search theory developed mainly in the field of labour supply functions is modified to suit a shopping destination context. In subsection 3.2 previous empirical results relating to this theory are outlined and random utility based logit models developed on the Adelaide data set to test implications of the theory.

3.1 A THEORY OF STORE CHOICE SET DETERMINATION

Classical economic consumer theory requires that each shopper possess perfect knowledge concerning the facilities destinations offer, their location, and the levels of transport services to them. This requirement, however, is in contradiction with evidence from psychology concerning the way information is stored in the brain and about the limited capacity of an individual to cope with information (Burnett 1973, Miller 1956). Psychologists contend that it is impossible for the human brain to comprehend total reality. There is then a constraint on the amount of information an individual can store. Given this constraint individuals must decide what information to store and what to discard. Economists would argue that this decision is based on factors associated with any optimal allocation problem; specifically, information will be stored whenever the marginal returns from so doing exceeds the costs in terms of money, time and effort.

Dealing with the benefits from search, first, it is reasonable to assume that as a passive recipient of information the individual will form a knowledge of the existence of destinations and the frequency distribution of the attractiveness of stores. In Chapter 2 the area delineated by this minimal knowledge holding was labelled 'awareness space'. The frequency distribution contains information on average store attractiveness and the dispersion around this, but not the attractiveness of individual stores.

A general formula for the maximum expected store attractiveness after searching N stores from a probability distribution of store attractiveness indices, F(A), is (David 1970, Rothschild 1974):

$$E(A^{*}|N) = \int_{0}^{\infty} \left[1 - F\ (A)\right]^{N} dA \qquad (5.6)$$

where $A^{*} = \max \left\{A_1, A_2, \ldots, A_N\right\}$ and A_i is an index of attractiveness at the ith store. $E(A^{*}|N)$ will be a decreasing function of N.

The marginal expected benefits from one more search can be calculated as:

$$\begin{aligned} f(N) &= E(A^{*}|N-1) - E(A^{*}|N) \\ &= \int_{0}^{\infty} \left[1 - F(A)\right]^{N-1} F(A)\ dA \end{aligned} \qquad (5.7)$$

which from Table 5.5, derived from David (1970), can also be seen to be a decreasing function of N. Where there is just one search the expected maximum value is the mean; it is equally probably that a consumer will have picked a very unattractive store as a very attractive store. As search increases to include two stores the expected maximum store attractiveness (for a normal distribution) is 0.5642 standard deviations above the mean, for three stores it increases to 0.8463 standard deviations above the mean, and so on. It is apparent from Table 5.5, however, that as N increases there is a very rapid reduction in the benefits from extra search. For instance, the expected marginal returns from increasing search from 5 stores to 20 stores is about equivalent to the (expected) extra benefits from searching just one other store when only one has been searched.

The pattern of expected benefits from search exhibited in Table 5.5 holds for any given level of store use. Generalizing this, when varying degrees of store use are admitted there will be a positive relationship between store use and search benefits. To provide a trivial example, suppose assessment of store attractiveness depended just on grocery prices, then those individuals who purchased the greatest quantities of groceries would also benefit most, in terms of reduced grocery expenditures, from discovering stores offering these products at low prices.

TABLE 5.5

THE EXPECTED VALUE OF THE MAXIMUM STORE ATTRACTIVENESS AFTER SAMPLING N STORES FROM A STANDARDISED PROBABILITY DISTRIBUTION

(mean = m, standard deviation = 1)

Number of Stores Searched (N)	Expected Maximum Store Attractiveness (any distribution)	Expected Maximum Store Attractiveness (normal distribution)
1	m	m
2	m + 0.5774	m + 0.5642
3	m + 0.8944	m + 0.8463
4	m + 1.1339	m + 1.0294
5	m + 1.3330	m + 1.1630
20	m + 3.0424	m + 1.8673

The concept can be extended , in a utilitarian spirit, to encompass other aspects of store attractiveness. Thus the total expected benefits from searching N stores can be expressed as:

$$TB_N = TB_N\left[(E(A^*|N), g)\right] \qquad (5.8)$$

where g is a measure of store use and $\partial g/\partial A^*$, the demand effect of discovering a more attractive store, is assumed to be negligible.

Equation (5.8) describes the benefits from search for once off store use situations (as, for instance, in purchasing an automobile). When stores are used in successive time periods (as for grocery purchases) benefits will accrue not only in the current period but also in future periods. Assuming a perfect correlation over time between the attractiveness of stores, the consumer will simply discount these benefits in calculating an optimal search policy and all search will occur in the initial period. If a less than perfect correlation exists between attractiveness measures over time, total discounted benefits from initial search will decrease and it will be profitable to also conduct search in subsequent periods (see Stigler 1961, p. 179).

Costs of search can be expressed as a function of opportunity time costs and direct expenditure on search. Given the search scenario portrayed above, the rational individual will systematically search shopping destinations at increasing distances from the home*. Looking at the variation in search costs across individuals, those with better than average access to shopping destinations will face lower search costs and, ceteris paribus, find it profitable to engage in more than average search activity.

---

* Strictly this statement holds only when all shopping journeys are of the type home-store-home. Once multi-trip journeys are recognized, the least search cost strategy cannot simply be determined, and other principal activity centres, such as the work place, may act as additional bases for conducting search.

Similar considerations apply to those with access to a car compared to individuals without such access.

From the above, two empirical models were developed to investigate the composition of store choice sets in the ATDATAS major grocery shopping subset. One model corresponds to the first term in the Burnett and Hanson (1982) formulation of the choice process as presented in equation (2.6). The concern of this model is with the probability, $P(j \in N_q)$, that store j forms an element in the choice set of individual q.

The second model was designed to examine more directly the determinants of store choice set size. From the theory presented it may be deduced that the summed amount of search activity for an individual, as observed in store choice set size, will be a function of the expected benefits to be obtained from search, principally measured by grocery shopping expenditure, and the expected costs of search, contributing factors being accessibility and ease of mobility.

Let the expected utility accruing to the individual from the benefits of searching m destinations be represented by:

$$V^B_{mq} = m\overline{V}^B_q (E_q, O^B_q) + \delta^B_m + \epsilon^B_{mq} \qquad (5.9)$$

where E_q is grocery shopping expenditure by individual q, O^B_q is a vector of other characteristics affecting utility obtained from the beneficial aspects of search, δ^B_m is a constant assumed to capture the expected benefits associated with the number (m) of destinations searched and ϵ^B_{mq} is included to represent the portion of expected 'search beneficial' utility hidden from the analyst.

In similar vein let the expected disutility accruing to individual q from the costs of searching m destinations be represented by:

$$V^{C}_{mq} = m\overline{V}^{C}_{q}\ (R_q,\ O^{C}_{q}) + \delta^{C}_{m} + \epsilon^{C}_{mq} \qquad (5.10)$$

where R_q is a vector of characteristics representing the accessibility of individual q to shopping destinations and other terms are the search cost counterparts of the search benefit terms explained for equation (5.9).

In the interests of notational economy equations (5.9) and (5.10) may be collapsed to

$$V^{BC}_{mq} = m\overline{V}^{BC}_{q}\ (E_q,\ R_q,\ O^{BC}_{q}) + \delta^{BC}_{m} + \epsilon^{BC}_{mq} \qquad (5.11)$$

where $\overline{V}^{BC}_{q}(E_q,\ R_q,\ O^{BC}_{q}) = \overline{V}^{B}_{q}(E_q,\ O^{B}_{q}) - \overline{V}^{C}_{q}\ (R_q,\ O^{C}_{q})$,

$\delta^{BC}_{m} = \delta^{B}_{m} - \delta^{C}_{m}$, $\epsilon^{BC}_{mq} = \epsilon^{B}_{mq} - \epsilon^{C}_{mq}$, and O^{BC}_{q} is a vector defined

as $O^{BC'}_{q} = (O^{B'}_{q},\ O^{C'}_{q})$. V^{BC}_{mq} may be thought of as the net utility associated with searching m destinations.

Individual q will find it profitable to search m, and no more than m destinations, if:

$$V^{BC}_{mq} > V^{BC}_{(m-1)q} \text{ and } V^{BC}_{(m+1)q} < V^{BC}_{mq}$$

or,

$$\delta^{BC}_{(m-1)} - \delta^{BC}_{m} + \epsilon^{BC}_{(m-1)q} - \epsilon^{BC}_{mq} < \overline{V}^{BC}_{q} < \delta^{BC}_{m} - \delta^{BC}_{(m+1)} + \epsilon^{BC}_{mq} - \epsilon^{BC}_{(m+1)q} \qquad (5.12)$$

Equation (5.12) represents an operational model for the term $P_q(C_m \in J_q)$, comprising one component of the DCS model presented in equation (2.8). Expanding equation (5.12) in terms of the notation in Chapter 2, yields:

$$C_m = 1 \text{ if } \quad -\infty < \bar{V}_q^{BC} < \delta_1^{BC} - \delta_2^{BC} + \epsilon_{1q}^{BC} - \epsilon_{2q}^{BC}$$

$$C_m = 2 \text{ if } \quad \delta_1^{BC} - \delta_2^{BC} + \epsilon_{1q}^{BC} - \epsilon_{2q}^{BC} < \bar{V}_q^{BC} < \delta_2^{BC} - \delta_3^{BC} + \epsilon_{2q}^{BC} - \epsilon_{3q}^{BC}$$

$$\vdots$$

$$C_m = N \text{ if } \quad \delta_{N-1}^{BC} - \delta_N^{BC} + \epsilon_{Nq}^{BC} - \epsilon_{(N+1)q}^{BC} < \bar{V}_q^{BC} < +\infty$$

If it is assumed that the $\epsilon_{mq}^{BC} - \epsilon_{(m+1)q}^{BC}$, etc. in equation (5.12) are iid extreme value type 1 then the probability of searching the (m+1)th destination, given that m destinations have been searched, can be described by a binary logit model:

$$\text{Prob}\left\{ V_{(m+1)q}^{BC} > V_{mq}^{BC} \mid V_{mq}^{BC} > V_{(m-1)q}^{BC} \right\} = P_{(m+1)q|mq}$$

$$= \frac{1}{1 + \exp\left(-\bar{V}_q^{BC} - \delta_{m+1}^{BC} + \delta_m^{BC}\right)} \qquad (5.13)$$

Further, the probability of searching m destinations may be expressed as a product of the independent binary probabilities:

$$\text{Prob}\left\{ V_{(m+1)q}^{BC} < V_{mq}^{BC} \mid V_{mq}^{BC} > V_{(m-1)q}^{BC} \right\} = P_m$$

$$= (1 - P_{(m+1)q|mq}) \prod_{k=1}^{m_q} P_{kq|(k-1)q} \qquad (5.14)$$

where m_q is the number of destinations searched by individual q. The log-likelihood function for a sample of Q individuals is:

$$\Gamma = \sum_{q=1}^{Q} \log(1 - P_{(i+1)q|iq}) \sum_{k=1}^{m_q} \log\left(P_{kq|(k-1)q}\right) \qquad (5.15)$$

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 'exploded' logit model of Beggs et al. (1981)* and the 'ordered' logit model described, for example in Maddala (1983, pp 46-49).

Examining the values of the terms $\delta_m^B - \delta_m^C$, it is to be noted that as m increases, from the discussion above on search costs, δ_m^C will also increase. In contrast, from Table 5.5, as m increases δ_m^B will decrease. The net effect is that as m increases δ_m^{BC} will decrease. Moreover, from previously presented information δ_m^{BC} can be expected to decrease at an increasing rate. As a result $\delta_{m+1}^{BC} - \delta_m^{BC}$ $(=\bar{\delta}_m^{BC})$ can be expected to decrease as m increases.

3.2 EMPIRICAL RESULTS ON STORE CHOICE SET DETERMINATION

Results from estimating the two models using the ATDATAS grocery shopping data are shown in Tables 5.6 and 5.7.

The form of the model used for $P(j \in N_q)$ was:

$$P_q(j \in N_q) = 1 \;/\; [1 + \exp\,(Z_q\,\alpha)] \qquad (5.16)$$

where Z_q is a row vector of independent variables not differentiated by alternative and α is a parameter vector. The dependent variable for this model took a value of 1 when destination j, from the objective choice set, also formed an element of the reported choice set, and 0 otherwise. A random selection of destinations from the objective choice set was used in estimating this model. Results from this model are shown in Table 5.6.

Three variables, GDISTANCE, LICENCE AND NCARS, were used to represent the accessibility of individuals to stores. As anticipated, distance to store took a negative sign and was the major determinant of whether a store from

---

* Beggs et al. dubb their model an ordered logit model. It is a generalisation of the ordered logit model described by Maddala and others. A more descriptive nomenclature for the model of Beggs et al. is 'exploded' logit.

TABLE 5.6

**STORE CHOICE SET DETERMINATION:
PROBABILITY STORE j LIES WITHIN
THE REPORTED CHOICE SET OF INDIVIDUAL q**

Variable Name*	Parameter Estimate	Standard Error	T-Statistic
GDISTANCE	-1.219	0.0602	-20.26
GSPENDTRIP	0.0043	0.0016	2.709
HINCOME	0.0020	0.0066	0.304
LICENCE	0.1880	0.1561	1.204
NCARS	0.1350	0.0778	1.735
RESYEARS	0.0017	0.0009	1.797
$(RESYEARS)^2$	-2.4×10^{-6}	1.5×10^{-6}	-1.587
SECEDUC	0.1934	0.1731	1.117
AL(1)	0.6017	0.2209	2.724
ρ^2	0.589		
% correctly predicted			
- at zero	50		
- at convergence	87		

* Note: Variables defined in Table 4.9.

stores. As anticipated, distance to store took a negative sign and was the major determinant of whether a store from the objective choice set formed an element in reported choice sets. Conversely the possession of a driver's licence and the number of cars in the household positively influenced the probability of a store lying within an individual's reported choice set.

The variable used to characterize the level of grocery shopping activity was 'GSPENDTRIP', the average expenditure on groceries on each major grocery shopping trip. The positive sign of the parameter estimate for this variable indicates that those with higher levels of grocery shopping expenditure have a greater probability of including any given store in their reported choice sets.

From theoretical considerations, the impact of income on search activity is ambiguous. On the one hand an increase in income increases the value of time and hence search costs. On the other, for normal goods, an increase in income will increase demand for the good and hence search benefits. In the event, from Table 5.6, household income (HINCOME) appeared to have little effect on store choice set determination.

The variables RESYEARS and $(\text{RESYEARS})^2$, being first and second power forms of the number of years of residency at the current home, were included in recognition of the 'stock of knowledge' aspect of store choice sets. It seems probable that, ceteris paribus, those who have lived in an area many years will be more likely to have searched any particular store and have included it in their choice sets. The positive sign of the RESYEARS parameter estimate empirically confirms this reasoning. The negative sign of the second power term suggests that store search activity declines as the period of residency increases.

The convenient linear specification was also invoked for $\overline{V}_q^B - \overline{V}_q^C$ ($= Z_q \alpha$) in the logit model of store choice set size detailed in equations (5.13) - (5.15). The maximum store choice set size considered by this model was '5+' (only three individuals specified six or more stores in their choice sets). With the exception of the variable GACCESS, the set of independent variables used was similar to the binary logit model of Table 5.6. GACCESS was designed to describe objectively the accessibility of individuals to grocery stores. It is defined as:

$$\text{GACCESS}_q = \sum_i (\text{GDISTANCE}_{iq})^{-2}$$

Various values for the power constant were tested. Results for a wide range of values did not alter the conclusions to be drawn from this analysis. Estimation results for the logit model of equations (5.13) - (5.15) are presented in Table 5.7. The model was estimated by first applying a method suggested by Winship and Mare (1984) and using the estimates so obtained as starting values in a Newton-Raphson evaluation of the log-likelihood functions shown in equation (5.15). The final estimates did not differ greatly from a regression analysis with the dependent variable being choice set size (and the parameters $\overline{\delta}_1^{BC}$, ..., $\overline{\delta}_4^{BC}$ omitted).

Again all parameter estimates took their postulated signs. As before, the accessibility variables (GACCESS, LICENCE and NCARS) were important determinants of reported choice set size, as was expenditure on grocery items (GSPENDTRIP). In contrast to the binary logit model, the years of residency variables were not even statistically significant at the 80% level. The last observation suggests a complex dynamic process at work in choice set determination. As new stores are searched they are not simply added to the existing choice set; rather the existing choice set is revised, with some stores being dropped from consideration. Thus overall choice set size exhibits little

TABLE 5.7

LOGIT MODEL OF CHOICE SET SIZE

Variable Name*	Parameter Estimate	Standard Error	T-Statistic
GACCESS	0.0353	0.0130	2.704
GSPENDTRIP	0.0050	0.0022	2.242
HINCOME	0.0054	0.0084	0.642
LICENCE	0.7669	0.1971	3.892
NCARS	0.4166	0.1056	3.946
RESYEARS	0.0006	0.0013	0.501
$(RESYEARS)^2$	-0.9×10^{-6}	0.2×10^{-5}	-0.458
$\overline{\delta}_1^{BC}$	0.0687	0.3095	0.223
$\overline{\delta}_2^{BC}$	-1.952	0.6112	-3.194
$\overline{\delta}_3^{BC}$	-3.658	0.7072	-5.169
$\overline{\delta}_4^{BC}$	-6.935	1.3878	-4.997
ρ^2	0.365		
% correctly predicted			
- at zero	20		
- at convergence	45		

* Note: Variables defined in Table 4.9.

systematic variation with years of residency; however stores at the margin of consideration for the entire sample tend to have been searched by those who have lived at the same residence for a number of years. Finally, the parameters $\bar{\delta}^{BC}_1, \ldots, \bar{\delta}^{BC}_4$ took the expected values relative to one another.

Overall, the empirical work, set out in this section, confirms earlier suspicions that reported choice sets represent an orderly subset of objective choice sets. Furthermore, the reported choice subsets conform with expectations developed from economic search theory. It is interesting to note that variables such as GSPENDTRIP and some socio-economic descriptors that are theoretically and statistically significant in explaining choice set composition, would be difficult to incorporate in a destination choice model estimated using unranked alternatives. As a consequence destination choice models, developed in isolation from a well specified choice set generation process, may not be able to adequately represent both the destination choice and choice set generation aspects. It has sometimes been argued that current choice models do adequately represent both choice and choice set generation aspects.

4. CONCLUSION

The points to emerge from this chapter are:

(1) that store choice set specification appears to exert a statistically significant influence on shopping destination choice model parameter estimates,

(2) this impact is in harmony with theoretical considerations,

(3) variations in reported store choice sets conform with expectations from economic theory, and

(4) it is doubtful, given the apparent complexity of choice set formation processes, that current choice models can satisfactorily represent both these and the choice act itself.

With this last assertion in mind, it is comforting to know that, given the correctness of the MNL form in representing destination choice, no bias will be introduced by confining attention to the reported choice set. Biased parameter estimates may result, however, by erroneously using a super set of those alternatives considered by the individual, unless the choice set formation process is fully accounted for in the model. The significance of variables such as GSPENDTRIP in explaining variations in reported choice sets suggests that current models do not fully account for both choice and choice set formation.

CHAPTER 6

MODELLING MICRO LEVEL LINKAGES BETWEEN FOOD SHOPPING ACTIVITIES

1. INTRODUCTION

A further aspect of choice set identification that was highlighted in Chapter 2 concerned the overall dimensions of behaviour to be incorporated in the model structure. In the context of modelling shopping destination choices, past research has tended to treat all aspects of food shopping homogeneously or has concentrated on just one category of shopping in isolation from other categories and other activities. The latter course has been the one adopted in the present study, with most attention being devoted to grocery shopping.

In this chapter, however, an effort is made to explore the inter-relationships between destination choices for three classes of food shopping, meat, major grocery and greengrocery shopping. In particular, a model of urban food shopping choices is developed which incorporates multi-purpose and multi-stop shopping possibilities. It is the way in which individuals seek to combine destination choices for the three classes of food shopping activity that is of primary interest to the model. One possibility is to do all types in the one store (e.g. a large supermarket) at the same time. Alternatively, an individual may choose to shop in different stores. In this case decisions are required on the arrangement of travel. Intermediate shopping patterns can also be constructed.

The simple MNL model developed in this chapter simultaneously accounts for choice of store for the three types of food shopping and associated travel arrangements.

Following, in Section 2, is an outline of the model specification, including justification of the approach adopted. Section 3 contains the results from model estimation.

2. OUTLINE OF AN APPROACH TO MODELLING URBAN FOOD SHOPPING BEHAVIOUR

Two propositions underlie the development of the food shopping model outlined in this chapter. These are:

(a) that the planning horizon for many activities, including food shopping, spans a longer period than one day and may vary between individuals, and

(b) the food shopping activities of most households can be effectively isolated from their other activities with the exception of work activities and within home activities.

Support for the former proposition is forthcoming from authors such as Damm and Lerman (1981). The following empirical evidence from Adelaide also serves to corroborate this proposition. Relevant results from an Australian Government survey (ABS 1982) on the food shopping patterns of approximately 350,000 Adelaide households are shown in Table 6.1. Only four per cent of those sampled stated that they shopped irregularly, with 79 per cent reporting major food shopping cyclical frequencies of weekly, fortnightly or monthly. Analysis of the data set used in the current study (results of which are also shown in Table 6.1) supported these figures, but in more detail, with 76 per cent of sampled households stating that they did all major types of food shopping on the same weekly, fortnightly or monthly cycle; 14 per cent displaying diverse arrangements with different cyclical frequency functions for one or more of

TABLE 6.1

SHOPPING FREQUENCIES FOR MAJOR FOOD ITEMS

	1981 ABS Survey		1980 Adelaide Survey	
	No.	%	No.	%
more than weekly	59,400	18	35	10
weekly	211,150	64	227	64
fortnightly	45,900	14	43	12
monthly	3,300	1	–	–
irregularly	12,600	4	–	–
different shopping frequencies for meat groceries and greengroceries	NA	NA	51	14

NA = not applicable

the major food shopping types. From this information it does not appear as though individuals plan their major food shopping travel on an ad hoc daily basis, but over a longer time span resulting in a regular cycle of shopping activities.

With respect to the second proposition many authors (CRA 1976, Hanson 1980, Hensher 1976, Oster 1978) have noted the dominant role played by the work and home nodes both in an organisational dimension and as major points of departure or destination for tours which include intermediate stops for other activities. For instance, Hanson (1980) has shown with Swedish data that 44 per cent of all stops at non-home places were made while on the work journey. Furthermore, initial analyses of the Adelaide data revealed that food shopping activities were not generally linked with other activities, apart from other food shopping activities, home activities and work. This observation held both for chosen shopping patterns and those shopping options viewed by households as possible alternatives to their usual arrangements. Less than ten per cent of households supplying shopping information reported linking other activities with either their chosen shopping pattern or those shopping patterns seen by them as possible alternatives to their usual arrangements.

These propositions hold two important implications for model development. Firstly, the appropriate unit for modelling should recognise the dependency of choices made within a shopping cycle rather than treat each food shopping choice as an independent event. Secondly, food shopping activities may be modelled in isolation from all non-home activities except work. The latter has the beneficial effect of simplifying model specification.

From the perspective sketched above the situation confronting the household with respect to food shopping can be viewed in terms of choosing a particular food shopping pattern (say sp) from a set of patterns (SP). Each element in SP is in turn compositely defined by a set (m, g, v, t, a, f) where m, g and v represent the meat, grocery and greengrocery shops in the pattern, t the travel modes used, a the tour arrangement and f the cyclical frequency of food shopping activities. Enlarging the random utility formulation of spatial shopping choices introduced in equation (2.2), the method used by individual q to evaluate alternative food shopping patterns can be written as:

$$V_{q(sp)} = \overline{V}_{q(sp)} (Z_{q(sp)}, , \alpha) + \epsilon_{(qsp)} \qquad (6.1)$$

with the probability that household q chooses alternative sp represented by:

$$\text{Prob}\left\{I_q = sp\right\} = \text{Prob}\left\{\epsilon_{q(sp)'} - \epsilon_{q(sp)} \leq \overline{V}_{q(sp)} - \overline{V}_{q(sp)'}\right\} \qquad (6.2)$$

Assuming that the random elements are iid Gumbel Type 1 extreme value distributed, the choice model will take the MNL form:

$$\text{Prob}\left\{I_q = sp \,|\, SP_q\right\} = \exp\,(\overline{V}_{(sp)q}) \Big/ \sum_{(sp)' \in SP_q} (\exp\,(\overline{V}_{(sp)'q})) \qquad (6.3)$$

where SP_q is the set of shopping patterns from which a choice is made by individual q.

The model specification of equations (6.1) - (6.3) is similar to that used by Adler and Ben Akiva (1979), reviewed in Chapter 2. In application, however, it differs in a number of important respects to the earlier work:

(a) it operates over a complete activity cycle rather than over an arbitrarily chosen time period,

(b) it incorporates single stop, multi-purpose travel possibilities, and

(c) the development context is at a more micro level than that considered by Adler and Ben Akiva.

For reasons of computational simplicity three further restrictions were placed on the MNL model specified in equation (6.3). Firstly attention was restricted to only those households operating on a weekly food shopping cycle, thus obviating the need to consider shopping frequency decisions. The complete model can then be seen as:

$$\text{Prob}\left\{I_q^f = f \mid F\right\} = \text{Prob}\left\{V_{qf} \geq V_{qf'}, \text{ for all } f' \neq f\right\} \quad (6.4)$$

$$\text{Prob}\left\{I_q^{sp} = sp \mid SP_f\right\} = \text{Prob}\left\{V_{qsp|f} \geq V_{q(sp)'|f} \text{ for all } (sp)' \neq sp\right\} \quad (6.5)$$

where all sp in equation (6.5) have been redefined to be compositely determined by (m, g, v, t, a) and SP_f represents the set of alternative shopping patterns available for frequency f. Analysis was confined to equation (6.5) and the case where f = weekly.

Secondly, households with linkage patterns involving work in their perceived choice sets were removed from consideration. Only 16 per cent of households in the Adelaide survey fell into this category. Tour arrangements for remaining households were therefore of the type depicted in Figure 6.1. Ignoring direction of travel there are five basic tour arrangements; specifically, linking all shopping outlets on the one tour, undertaking three separate two trip tours, and various combinations of linking two outlets on the one tour and the third outlet on another tour. It should also be recognised that some of the trips depicted in Fig. 6.1 may be empirically insignificant short walk trips. For example, for tour arrangements of type 1, if G, M and V are located within the same supermarket, the only

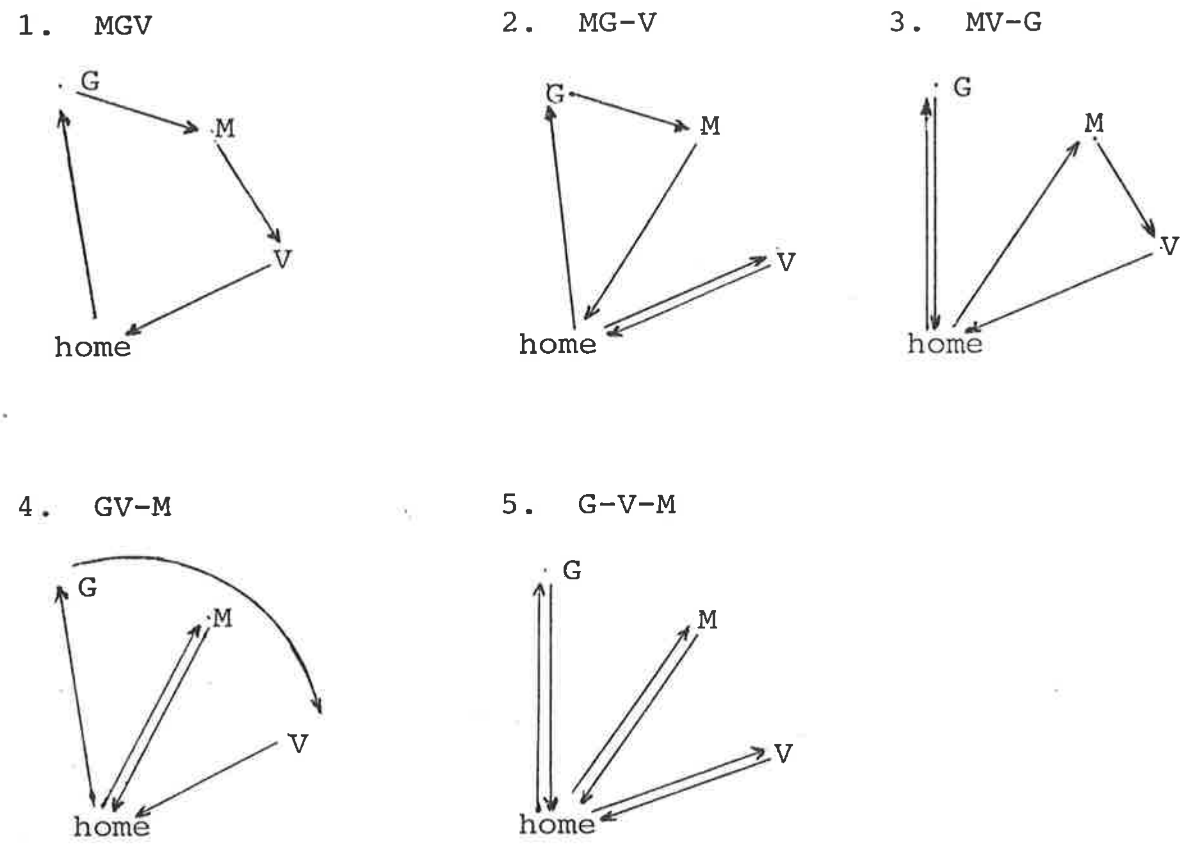


FIGURE 6.1: POSSIBLE FOOD SHOPPING ARRANGEMENTS

Notes:
1. M, G and V refer to meat, grocery and greengrocery outlets respectively.
2. Arrows indicate trips (sequencing of activities and direction of travel, examples only).

significant travel to occur is between home and the supermarket.

Finally, in common with empirical work presented in Chapters 4 and 5, a linear functional form of V_{sp} is assumed.

3. ANALYSIS AND MODEL ESTIMATION

The analysis subset used in the model estimation work reported in this section comprised 222 households. Restrictions placed on the full data set to arrive at the analysis data set were:

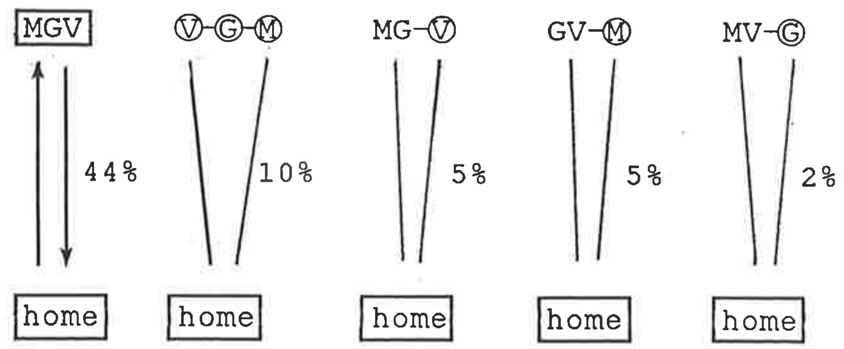
(a) as stated below equation (6.5), only households with weekly shopping patterns were admitted, and

(b) only those respondents were included who supplied information for meat, major grocery and green grocery shopping with home as the sole connecting activity both for chosen and non-chosen stores.

As explained above, these restrictions were utilized in the interest of simplification.

A pictorial representation of the main tour and shopping locational arrangements used by the 222 households comprising the analysis data set is displayed in Figure 6.2. As can be seen from Figure 6.2 almost half of the sample conducted all food shopping on a single stop multi-purpose tour. Of the remaining households about half completed all food shopping types on a single tour, but with multiple stops. The final 25 per cent of households displayed a variety of shopping arrangements involving two or more tours (often conducted on different days). The rich variety of food shopping patterns as displayed in this Figure contrasts sharply with the simplistic modelling approach used in the past to represent food shopping

Single Tour Arrangements



Multiple Tour Arrangements

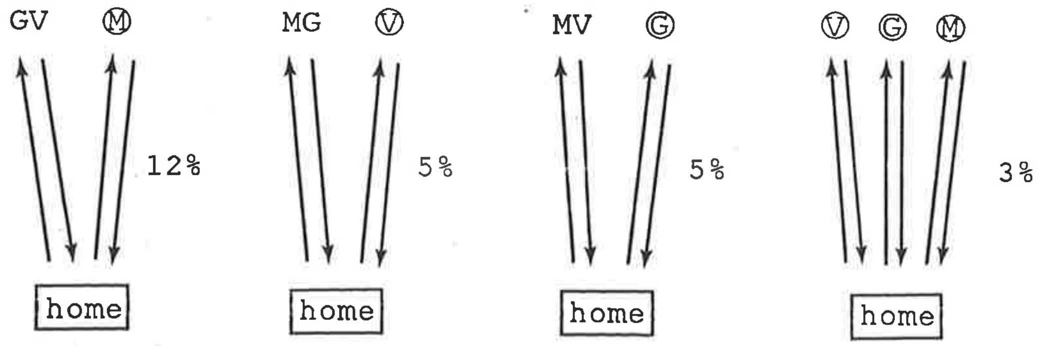


FIGURE 6.2 Classification of major selected food shopping travel arrangements by sampled households.

behaviour. Traditionally, these models have ignored individual participation in multiple shopping activities at the one location and travel linkage arrangements.

In addition to the shopping patterns actually selected by the sampled households as described in the previous paragraph, the MNL model of equation (6.3) requires the generation of additional shopping patterns that are seen by each household as viable options to the selected pattern. These were constructed using information supplied by respondents on the outlets they usually patronised for major meat, grocery and green grocery shopping, available methods of travel to these stores, and other stores and methods of travel that the household reported as possible options to the usual store and mode for each shopping type.

An example output of information supplied on these matters was displayed in Figure 4.2. Designating the meat stores in Figure 4.2 as M_1, M_2 and M_3, the grocery stores as G_1, G_2 and G_3, the fruit and vegetable stores as V_1 and V_2 and the modes, car and walk respectively as C and W, then the information in Figure 4.2 can be represented by these symbols as in the upper portion of Table 6.2 The available shopping patterns for the household are then found by taking all possible combinations of these stores and modes and overlaying on each combination the five possible tour types of Figure 4.1 as is done in the bottom half of Table 6.2. Certain shopping patterns were, however, disallowed according to one or more of the following criteria:

(a) no change of mode was allowed on different trips of the same tour except for short walk trips within a store or shopping centre (which were ignored), and

(b) walking tours of more than 60 minutes duration were omitted.

It should also be noted that even for the simple example of Figure 4.2 the number of alternative shopping patterns

TABLE 6.2

GENERATION OF FOOD SHOPPING PATTERN ALTERNATIVES
(FROM THE EXAMPLE DISPLAYED IN FIG 2)

1. Supplied Data

Meat Data		Grocery Data		Fruit + Vegetable Data	
Store	Mode	Store	Mode	Store	Mode
M_1	C	G_1	C	V_1	C
M_1	W	G_1	W	V_1	W
M_2	C	G_2	C	V_2	C
M_2	W	G_2	W	V_2	W
M_3	C	G_3	C		
M_3	W				

2. Generated Shopping Patterns

Store and Mode Combinations						Permissible Tour Arrangements				
Meat Store	Meat Mode	Grocery Store	Grocery Mode	Fruit Store	Fruit Mode	MGV	MG-V	MV-G	GV-M	G-V-M
M_1	C	G_1	C	V_1	C	✓	✓	✓	✓	✓
M_1	C	G_1	C	V_1	W		✓			✓
M_1	C	G_1	C	V_2	C	✓	✓	✓	✓	✓
M_1	C	G_1	C	V_2	W		✓			
M_1	C	G_1	W	V_1	C			✓		✓
M_1	C	G_1	W	V_1	W					✓
.	.	.	.	.	.				✓	✓
.	.	.	.	.	.					
.	.	.	.	.	.					
.	.	.	.	.	.					
M_3	W	G_3	C	V_2	W			✓		✓

D

generated is quite large (330 to be exact). Consequently it would have been impracticable to use complete choice sets in estimating the MNL model. Fortunately, as was brought to notice in the past chapter, complete choice sets are not needed in order to obtain consistent parameter estimates from an MNL model and estimation was achieved using random subsets of alternatives (which included chosen options).

The specification used in model estimation for the conditional indirect utility functions was:

$$\begin{aligned}
\overline{V}_{(sp)q} = \; & \alpha_1 \log(\mathrm{MPRICE}_{(sp)q}) + \alpha_2 \log(\mathrm{GPRICE}_{(sp)q}) \\
& + \alpha_3 \log(\mathrm{VPRICE}_{(sp)q}) + \alpha_4 \mathrm{MQUAL}_{(sp)q} \\
& + \alpha_5 \mathrm{GSEL}_{(sp)q} + \alpha_6 \mathrm{VQUAL}_{(sp)q} + \alpha_7 \mathrm{MCONV}_{(sp)q} \\
& + \alpha_8 \mathrm{GCONV}_{(sp)q} + \alpha_9 \mathrm{VCONV}_{(sp)q} \\
& + \alpha_{10}(T - t_{(sp)q}) + \alpha_{11}(\mathrm{HINCOME}_q - c_{(sp)q}) \\
& + \alpha_{12} \mathrm{NUMBERLOCS}_{(sp)q} + \alpha_{13} \mathrm{NUMBERTOURS}_{(sp)q} \\
& + \alpha_{14} \mathrm{CARSAV}_{(sp)q} + \alpha_{15} \mathrm{LIFECYCLEYC}_{(sp)q}
\end{aligned}$$

Two variables were included to represent travel cost effects, namely, TTIMEN and TCOSTN. Again it was necessary to use network measures because, although perceived travel time and cost information were available between the individual's home and stores contained in the shopping patterns, such information was not available for travel between spatial separate stores contained in the shopping pattern. For instance, for shopping pattern tour arrangement 1 of Figure 6.1 perceived travel time and cost information was available for the trips home $\rightarrow$ G and V $\rightarrow$ home, but not for trips G $\rightarrow$ M and M $\rightarrow$ V. The total travel time for any shopping pattern was obtained by simply adding the travel times for each tour. A similar construction was used for total travel costs.

Two variables (NUMBERLOCS and NUMBERTOURS) were included to measure the effect on utility of differences in tour arrangements. The inclusion of these variables follows the arguments of Adler and Ben Akiva (1979) concerning the scheduling convenience associated with an activity pattern (see Chapter 2).

The attractiveness of stores visited in the travel pattern was represented by a set of nine variables, comprising the three indices of PRICE, SELECTION/QUALITY, and STORE CONVENIENCE for each of the meat, major grocery, and green grocery activity categories conducted on the travel pattern.

A final group of variables used in the model relates to socio-economic characteristics of the household. The variable measuring car availability, CARSAV, was designed to capture competition for cars within a household. The final variable 'LIFECYCLE' demonstrates the capability of the model to incorporate the effect of socio-economic differences on tour arrangements. This variable was formulated so as to encapsulate the time constrained lifestyle of households with young children (as demonstrated for instance, by research into activity patterns at the Transport Studies Unit, Oxford; Jones <u>et al</u> 1980).

Results from model estimation are displayed in Table 6.3. Most variables took on their postulated signs and were significant at the 95 per cent level. The major exception to this was the variable measuring the number of locations visited in the shopping pattern. From the reasoning of Adler and Ben Akiva it was expected that the parameter of this variable would be positive. That this variable took on a negative parameter estimate of high statistical significance, however, is not altogether surprising. In particular, the earlier reasoning relied to a large degree on the inability of the shopping outlet attractiveness

TABLE 6.3

ESTIMATION RESULTS: SHOPPING TRAVEL PATTERN MODEL

Variable Name*	Parameter Estimate	Standard Error	T-Statistic
TTIMEN	-0.0443	0.0110	-4.036
TCOSTN	-0.0132	0.0082	-1.612
NUMBERLOCS	-1.090	0.2097	-5.197
NUMBERTOURS	0.6970	0.1092	6.383
log(MPRICE)	-1.978	0.6444	-3.070
log(GPRICE)	-1.544	0.6201	-2.489
log(VPRICE)	-0.7296	0.6692	-1.090
MQUALITY	1.179	0.2211	5.332
GSEL	0.9814	0.2008	4.888
VQUALITY	0.8362	0.2251	3.714
MCONV	0.2487	0.1853	1.342
GCONV	0.5339	0.1701	3.139
VCONV	0.0160	0.2416	0.066
CARSAV	1.185	0.426	2.781
LIFECYCLEYC	1.453	1.081	1.345

ρ^2	0.49
% correctly predicted	
- at zero	10
- at convergence	63

*Note: Variables defined in Table 4.9

variables to detect adequately differences in utility between outlet alternatives. Evidently these are being measured with reasonable accuracy. Apparently what is not being totally captured elsewhere in the model (specifically, by the accessibility related variables) is the convenience of shopping at just one location, (i.e. within the one store or shopping centre) and this is reflected in the negative coefficient for this variable.

Examining the remaining variables, it will be noted that the parameter estimates attached to the accessibility related variables ('TTIMEN' and 'TCOSTN') are negative, signifying a desire by households to minimise food shopping travel. As expected the parameter estimate of the variable 'NUMBERTOURS' is positive, expressing the addition to utility of being able to schedule activity needs as they arise.

Concerning the nine shopping outlet 'attractiveness' related variables, negative relationships were found between perceived high prices and choice of outlets. Conversely as ratings of selection, quality and convenience increased, so did the probability of selecting those outlets. It is interesting to note that the relative magnitude, between food shopping categories, of the parameter estimates attached to these outlet attractiveness variables closely mirror differences in reported expenditures on meat, groceries and green groceries. This result is not surprising, given the relationship between store choice and shopping expenditure revealed in Chapter 3.

The remaining two variables measure differences in shopping pattern utilities due to variations in the socio-economic makeup of households. Conforming to a priori expectations the parameter estimate of the 'CARSAV' variable was found to be positive. Also as anticipated, the parameter estimate of the variable 'LIFECYCLE' is positive, indicating a preference by those households where all children are less than five years old, to complete all food shopping activities on the one tour. This preference may reflect a squeezed time schedule after discharging child care duties, as previously hypothesised, or a desire to limit shopping activities when young children are present.

To gauge model performance, the model of Table 6.3 was compared against the simultaneous MNL models of mode/store choice presented in Chapter 4, re-estimated using the reduced data set. Summary statistics of store choice prediction for the model of Table 6.3 ('tour' model) and the separately estimated models were then derived. These are shown in Table 6.4 Overall goodness of fit for the four models can be assessed by reference to the respective ρ_c^2 statistics. It is evident from these statistics that the 'tour' model fits the data considerably better than the separately estimated shopping type models. The other two measures shown (percent correctly predicted and summed residuals) relate more particularly to prediction of store choice. Surprisingly there is little to discriminate between the 'tour' model and respective separately estimated food shopping type models on the basis of percent of store choices correctly predicted. The summed residuals, however, are 10% - 15% higher for the separately estimated models than for the 'tour' model. Because, for any shopping type, the same set of individuals and stores were used in the tour model as in the separately estimated models, the absolute level of the residuals are directly comparable. It is apparent from these statistics as a whole, that the 'tour' model provides a better basis for describing and understanding food shopping choices than the separately estimated models.

TABLE 6.4

COMPARISONS BETWEEN TOUR MODEL AND SEPARATELY ESTIMATED MODELS FOR EACH TYPE OF FOOD SHOPPING

Model	ρ^2	Meat Shopping		Grocery Shopping		Vegetable Shopping	
		% Store Choices Correctly Predicted	Summed Store Choice Residuals	% Store Choices Correctly Predicted	Summed Store Choice Residuals	% Store Choices Correctly Predicted	Summed Store Choice Residuals
'tour' model	.49	78.1	203.7	74.5	225.2	76.2	170.3
separate meat shopping choice model	.29	70.2	238.8				
separate grocery shopping choice model	.31			63.4	266.3		
separate vegetable shopping choice model	.27					68.7	193.1

Note: Summed residuals are calculated as:

$$RESID = \sum_{q=1}^{Q} \sum_{i=1}^{N_{qs}} \left| k_{qis} - \hat{P}_{qis} \right|$$

where k_{qis} is equal to 1 if individual q chooses store i for shopping type s and 0 otherwise

P_{qis} is the predicted probability that individual q chooses store i for shopping type

Q is the set of individuals

N_{qs} is the set of stores of shopping type s for individual q

4. CONCLUSION

This chapter has represented a first step in the analysis of shopping linkages at a micro level. A model was constructed which permitted a greater range of food shopping possibilities than had been incorporated in past shopping behaviour research, including much at the research appearing under the current study. Parameter estimates from this model were shown to conform to intuitively plausible constructs of food shopping behaviour. Furthermore, the model more closely replicated actual store choices than models developed separately for each food shopping type.

This line of development, however, is not further pursued empirically in this study. Indeed, as was noted in Chapter 2, there are problems associated with the model methodology utilized in this chapter, particularly when further behavioural complexities are admitted. The results presented in this chapter serve only to indicate that linkages between food shopping categories may be important. In the following chapter another area of shopping behaviour is explored, namely, the inter-relationship between store and shopping expenditure choices. In the final chapter a method is presented which potentially accounts both for linkages between classes of shopping activity and between shopping and participation in other activities, and for the inter-relationship between destination and activity intensity choices.

CHAPTER 7

AN INTEGRATED EMPIRICAL MODEL OF GROCERY SHOPPING STORE CHOICE AND LEVEL OF EXPENDITURE

1. INTRODUCTION

A major focus of the theoretical framework presented in Chapter 3 was the inter-relationship between shopping destination and expenditure choices. Notably, it was shown how a particular assumed form for the conditional indirect utility functions used in the destination choice model, given the constructs of economic theory, logically leads to a particular form for the shopping expenditure functions. In this chapter estimation results from an integrated model, embracing jointly shopping destination and expenditure decisions are presented. Data for this exercise were derived from the grocery section of the shopping questionnaire and grocery shopping expenditures gleaned from the activity diaries.

A specification, considered in Chapter 3, for the conditional indirect utility functions was:

$$V_{iq} = \left[\alpha_3 - \alpha_2 \log(p_i/\psi_{iq}) + \alpha_1 (Y_q - c_{iq}) + \alpha_4 (T - t_{iq})\right] (p_i/\psi_{iq})^{\alpha_5} + \epsilon_{iq} \qquad (7.1)$$

which, after application of Roy's identity, was shown to yield a shopping expenditure model:

$$E_{iq} = \frac{\alpha_5 \alpha_3}{\alpha_1} + \frac{\alpha_2}{\alpha_1} - \frac{\alpha_2 \alpha_5}{\alpha_1} \log(p_i/\psi_{iq}) + \alpha_5 (Y_q - c_{iq}) + \frac{\alpha_5 \alpha_4}{\alpha_1} (T - t_{iq}) + u_{iq} \qquad (7.2)$$

where the ϵ_{iq} and u_{iq} are error terms included to represent choice influencing factors hidden from the analyst. Because, however, equations (7.1) and (7.2) form an integrated system, with both sets of error terms arising from the same source, namely analyst uncertainty concerning the conditional indirect utility functions, on estimation explicit allowance should be made for potential correlation between the ϵ_{iq} and u_{iq}.

A further complication is that shopping expenditure is only observed at chosen stores. Optimal expenditure levels for each individual at non chosen stores are concealed from the analyst. These features firmly place equation system (7.1) and (7.2) in the realm of sample selectivity models.

This chapter is developed as follows. Section 2 contains a review of sample selection models applicable to the estimation of equation system (7.1) - (7.2). To begin with, the review is simplified by assuming that equation (7.1) is characterised by a binary choice. Once this has been covered, the review is extended to cover polychotomous choice contexts. In Section 3, sample selection models are compared with mixture models and reasons enunciated why the former model genre was preferred in the current work. Section 4 is then devoted to estimating an integrated store choice/shopping expenditure model. In the first part of this section certain theoretical and consistency checks are developed. In the latter part, results from model estimation are presented.

2. STATISTICAL ISSUES OF SAMPLE SELECTIVITY

For reasons of notational economy, the discrete/continuous choice model system, implied by equations (7.1) and (7.2), is recouched here as:

$$V_{iq} = \overline{V}_{iq} (Z_{iq}, \alpha) + \epsilon_{iq} \qquad (7.3)$$

$$E_{iq} = X_{iq}\beta + u_{iq} \qquad (7.4)$$

where V_{iq} is unobservable, but has an observable polychotomous realization, such that I_q = j if $V_{jq} > \text{Max } V_{iq}$ (i = 1, 2, ..., N, i≠j), and E_{jt} is only observed if I_q = j. In this section two methods that are capable of accommodating aspects of the dependency between equations (7.3) and (7.4) are outlined. The major focus is on the Heckman-Lee (H-L) sample selectivity correction method. An alternative method developed by Hay/Dubin and McFadden (H-D+M) is also briefly discussed.

2.1 THE HECKMAN/LEE SELECTIVITY CORRECTION FACTOR

To explain the H-L method assume first of all that the realization of V_{iq} is dichotomous (N=2); that is, only two stores are available in which to shop. With this simplification an unobservable variable w_q may be defined:

$$\begin{aligned} w_q &= V_{1q} - V_{2q} + \epsilon_{1q} - \epsilon_{2q} \\ &= \tau(\alpha, Z_q) + \epsilon^*_q \qquad (7.5) \end{aligned}$$

where $\tau(\alpha, Z_q) = V_{1q} - V_{2q}$ and $\epsilon^*_q = \epsilon_{1q} - \epsilon_{2q}$. Since the scale of w_q is arbitrary it is convenient to specify a zero threshold value and normalize the variance of ϵ^*_q to 1. Thus, if $w_q > 0$ expenditure is observed at store 1, otherwise expenditure is observed at store 2.

The interpretation of a variable such as w_q need not be restricted to the store utility based analysis presented above. Other examples of w_q are propensity to participate in a survey or government program (e.g. Barnow et al. 1981), join a union (e.g. Lee 1978), and seek college education (e.g. Kenny et al. 1979). More generally, at least three types of sample selectivity may be identified. Samples may be non-random due to individual decisions (e.g. on whether or not to participate in a survey) yielding the observed data points, termed self-selection, due to decisions of a sampling administrator, termed administrator selection, or because of individuals falling out of a panel survey despite an initial random sample, termed attrition selection.

Following through the shopping behaviour analysis, a case of self-selection, Figure 7.1 illustrates for an X_{iq} vector of length two, one of which is a constant, the likely situation where w_q is positively correlated with E_{1q}, the latter variable representing shopping expenditure at store 1. Circled observations lying within the population of interest (i.e. all shopping expenditures) are nevertheless excluded from the sample drawn from store 1 due to self selection. It can be seen that the application of a simple regression model, such as that shown in the equation of Figure 7.1, using OLS, will yield biased estimates of β_0 and β_1 since the error term u_q will be correlated with x_q. From Figure 7.1, for low values of x_q there is a tendency for u_q to be smaller than for large values of x_q. Ignoring selection will result in biased OLS estimates; however, it is difficult to determine whether the biased OLS estimates will understate or overstate the true causal effects (Berk 1983).

Mathematically the regression equation for the observed data points is:

$$E(E_{1q}|w_q > 0) = X_q\beta + E(u_{1q}|w_q > 0)$$

or,

$$E(E_{1q}|w_q > 0) = X_q\beta + E(u_{1q}|\ \epsilon_q^* > -\tau(\alpha, Z_q)) \qquad (7.6)$$

Also, u_{1q} may be regressed on ϵ_q^*:

$$u_{1q} = \varphi\,\epsilon_q^* + v_{1q} \qquad (7.7)$$

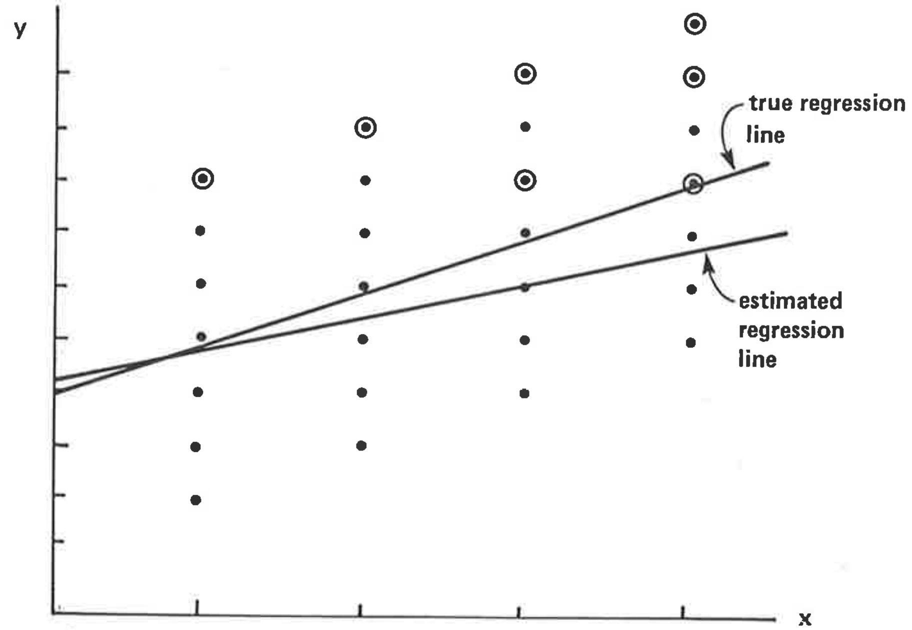


FIGURE 7.1 : ILLUSTRATION OF ESTIMATED AND TRUE REGRESSION LINES, $y = \beta_0 + \beta_1 x$, WHEN INCLUSION OF SAMPLE POINTS DEPENDS ON THE VALUE OF A SELECTION VARIABLE

where v_{1q} is a new error term uncorrelated with ϵ_q^* and $\varphi = \sigma_{u_{1q}\epsilon_q^*}/(\sigma_{\epsilon_q^*\epsilon_q^*})^2$ is the regression coefficient with $\sigma_{u_{1q}\epsilon_q^*}$ representing the covariance between u_q and ϵ_q^* and $(\sigma_{\epsilon_q^*\epsilon_q^*})^2$ the variance of ϵ_q^* (= 1). Using equations (7.6) and (7.7)

$$E(E_{1q}|w_q > 0) = X_{1q}\beta + \sigma_{u_{1q}\epsilon_q^*} E(\epsilon_q^*|\epsilon_q^* > -\tau(\alpha, Z_q)) \qquad (7.8)$$

If the ϵ_q^* and u_{1q} are normally distributed, the last term on the RHS of equation (7.8) can be shown to equal (Johnson and Kotz, 1972, pp 112-113) $-\phi[\tau(\alpha, Z_q)]/\Phi[\tau(\alpha, Z_q)]$ so:

$$E(E_{1q}|w_q > 0) = X_{1q}\beta - \sigma_{u_{1q}\epsilon_q^*}\frac{\phi[\tau(\alpha, Z_q)]}{\Phi[\tau(\alpha, Z_q)]} \qquad (7.9)$$

where Φ is the cummulative distribution function and ϕ the density function of the standard normal distributon.

Other distributional assumptions can be made about ϵ_q^* and u_{1q}, such as they are logistically distributed, and models analogous to equation (7.9) constructed (Muthen and Joreskog 1983). However, whereas the linear relationship of equation (7.7) follows automatically when ϵ_q^* and u_{1q} are bivariate normal, for other distributional assumptions it must be assumed.

Heckman (1976, 1979) identified a two stage procedure for obtaining estimates of the parameters contained in equation (7.9). The estimation sequence is:

1. estimate equation (7.5) using binary probit,

2. use $\hat{\alpha}$ and Z_q to calculate $\phi[\tau(\hat{\alpha}, Z_q)] / \Phi[\tau(\hat{\alpha}, Z_q)]$, and finally,

3. estimate equation (7.9) by OLS replacing $\phi[\tau(\alpha, Z_q)] \div \Phi[\tau(\alpha, Z_q)]$ with $\phi[\tau(\hat{\alpha}, Z_q)] / \Phi[\tau(\hat{\alpha}, Z_q)]$

Stage 1 involves use of the entire sample and stage 2 the subsample choosing store 1. This procedure provides consistent estimates of α, β and $\sigma_{u_{1q}\epsilon^*_q}$. Similar methods can be applied to estimate expenditure at store 2.

In extending the Heckman-Lee binary model to a multinominal choice situation, two problems are encountered. Firstly, the choice model can no longer be specified using just one error term. Secondly, the probit model is computationally intractible beyond about 3 alternatives. Lee (1983) has devised a method of circumventing these problems by (i) concentrating on the error term associated with the chosen alternative, and (ii) assuming that the ϵ_{iq} are iid extreme value type 1, so that the choice model is of the MNL form, but then transforming the error terms to a standard normal distribution.

Consider the full model of equations (7.3) and (7.4). Let,

$$\eta_{jq} = \text{Max } V_{iq} - \epsilon_{jq} \quad (i = 1, 2, \ldots, N, \; i \neq j) \tag{7.10}$$

It follows that:

$$I_q = j \text{ iff } \eta_{jq} < \overline{V}(Z_{jq}, \alpha) \tag{7.11}$$

This effectively reduces the choice of the jth alternative to a binary decision with mutual exclusivity; either it will be chosen or not. Assuming the ϵ_{jq}s are iid extreme value type 1 the probability of alternative j being chosen is given by a logit model. Also the distribution of η_{jq} is:

$$D(\eta_{jq}) = \exp(\eta_{jq}) / \left(\exp(\eta_{jq}) + \sum_{\substack{i=1 \\ i \neq j}} \left(\exp \; \bar{V}(Z_{jq}, \alpha) \right) \right) \quad (7.12)$$

To use the results from Johnson and Kotz, η_{jq} has to be transformed into standard normal. To achieve this Lee (1982 and 1983) drew attention to a well known transformation in statistics:

$$\eta^*_{jq} = J(\eta_{jq}) = \Phi^{-1}(D(\eta_{jq})) \quad (7.13)$$

where D generally can be any distribution function, but in the logit model under consideration is given by equation (7.12). The transformed error term η^*_{jq} will be a standard normal variate. Computationally simple and accurate methods for aproximating the inverse of the standard normal distribution can be found in the work of Bock and Jones (1968) and Hildebrand (1956). Errors of approximation for these methods are less than 3×10^{-4}. With this transformation j will be chosen iff $\eta^*_{jq} < J\,[\bar{V}(Z_{jq}, \alpha)]$.

Assuming u_{iq} is also normally distributed then the bivariate distribution between η^*_j and u_j can be specified as $N(0, 0, 1, \sigma_{u_j u_j}, \rho_{\eta^*_j u_j})$. The equation system (7.3) and (7.4) should only be estimated independently when the correlation coefficient, $\rho_{\eta^*_j u_j}$, is equal to zero. In the more general case, the conditional expectation $E(u_{jq} | I_q = j)$ needs to be included as a regressor in equation (7.4). But given the transformation of η, the binary H-L model can simply be generalised to derive the multinominal equivalent to equation (7.9) as*:

* Note that $\sigma_{u_j \eta^*_j} = \rho_{u_j u_j} \sigma_{\eta^*_j u_j}$ since $(\sigma_{\eta^*_j \eta^*_j})^2$ is equal to 1.

$$E_{jq} = X_{jq}\beta - \sigma_{u_j \eta_j^*} \phi \left\{ J[\bar{V}(Z_{jq}, \alpha)] \right\} / D[\bar{V}(Z_{jq}, \alpha)] + v_{jq} \tag{7.14}$$

with $E(v_{jq} | I_q = j) = 0$. The two-stage estimation procedure is:

1. estimate the logit model implied by equation (7.3), obtaining values for $\hat{\alpha}$,

2. For the chosen store calculate $J[\bar{V}(Z_{jq}, \hat{\alpha})] = \Phi^{-1}\left\{D[\bar{V}(Z_{jq}, \hat{\alpha})]\right\} = \Phi^{-1}(\hat{P}_{jq})$ and $D[\bar{V}(Z_{jq}, \hat{\alpha})] = \hat{P}_{jq}$,

3. estimate equation (7.14) using OLS with $\sigma_{u_j \eta_j^*}$ being the parameter estimate for selectivity correction, and

4. correct the variance/covariance matrices associated with the OLS estimation of equation (7.14). This correction is necessary because the v_{jq} are heteroskedastic. Correction formulas are derived in Appendix 7A.

Alternatively the system may be estimated using full information maximum likelihood. The log-likelihood function is:

$$\Gamma = \sum_q \sum_i \Bigg\{ k_{iq} \log\left\{ (\sigma_{u_i u_i})^{-1} \phi[(E_{iq} - X_{iq}\beta) \div \sigma_{u_i u_i}] \right\} + \bar{k}_{iq} \log \Phi \left\{ \left[J(Z_{qj}, \alpha) - \rho_{\eta_i^* u_i} [(E_{iq} - X_{iq}\beta)/\sigma_{u_i u_i}] \right] / [1 - (\rho_{\eta_i^* u_i})^2]^{1/2} \right\} \Bigg\} \tag{7.15}$$

where $J(Z_{qi}, \alpha) = J[\bar{V}(Z_{qi}, \alpha)]$ and $k_{iq} = 1$ if $I_q = i$ and 0 otherwise. Equation (7.15), however, may be difficult to evaluate.*

Finally, the exposition of the Heckman-Lee method has been in the context of ranked data. With this class of data it is possible to estimate separate expenditure models of the type shown in equation (7.14) for each chosen store, j = 1, 2, ..., N, and thus derive a set of selectivity correction factors. When unranked data is used to estimate the MNL model and only one expenditure equation is estimated it is necessary to assume $\sigma_{u_j \eta_j^*}$ is constant for all j.

2.2 THE HAY-DUBIN AND McFADDEN SELECTIVITY CORRECTION FACTOR

To derive the H-D+M method first define:**

$$\omega_{ji} = \epsilon_i - \epsilon_j \text{ and } \tau_{ji} = \overline{V}(Z_i, \alpha) - \overline{V}(Z_j, \alpha) \qquad (7.16)$$

so that I = j iff $\omega_{ji} < \tau_{ji}$. If the ϵ_i are distributed iid extreme value type 1 then the N - 1 random variables will be jointly distributed with a c.d.f.:

$$D(\omega_{j1}, \omega_{j2}, \ldots, \omega_{j,j-1}, \omega_{j,j+1} \ldots, \omega_{jN})$$

$$\equiv \overline{D}(\omega_j) = 1 / [1 + \sum_{\substack{k=1 \\ k \neq j}} \exp(-\omega_{jk})] \qquad (7.17)$$

By assumption:

$$E(\omega_{ji}) = 0, \; var(\omega_{ji}) = \pi^2/6 \text{ and } cov(\omega_j, \omega_i) = \Sigma_{\omega_i} \text{ for}$$

$i = 1, 2, \ldots\ldots, N, \; i \neq j.$

---

* In a binary choice context McFadden et al. (1986) have utilised a FIML approach. Mannering and Hensher (1986), however, in a recent review concluded that; 'Unfortunately, full information modelling systems are computationally cumbersome, and extensions beyond simple binary discrete choices do not appear promising. Indeed binary choice applications are only relatively straight forward if one uses a standard probit model with a linear specification of the indirect utility expression'.

** In the interests of notational economy the q subscript has been dropped in this section.

Generalising a result from Olsen (1980), if the conditional expectation of u_j given ω_j is linear, then:

$$u_j = \frac{\sigma_{u_j \omega_j}}{(\sigma_{\omega_j \omega_j})^2} \left[\omega_j - E(\omega_j) \right] + v_j$$

$$= \sum_{\substack{i=1 \\ i \neq j}} \left[\frac{6}{\pi^2} \rho_{\omega_i u_j} \sigma_{u_j u_j} \right] \omega_{ji} + v_j \qquad (7.18)$$

The continuous choice model can then be written as:

$$E(E_j | I = j) = X_j \beta + \sum_{\substack{i=1 \\ i \neq j}}^{N} \left[\frac{6}{\pi^2} \rho_{\omega_i u_j} \sigma_{u_j u_j} \right] E(\omega_{ji} | \omega_{j\ell} < \tau_{j\ell},$$

$$\ell = 1, 2, \ldots, N, \ \ell \neq j) \qquad (7.19)$$

As can be seen equation (7.19) is similar to equation (7.14). The H-D+M method, however, includes terms accounting for correlation between the error term in the demand model for the chosen alternative and error terms in the discrete choice model for non-chosen alternatives. In contrast the H-L method only includes one selectivity correction term in each demand model. This term accounts for correlation between the demand model error term and the error term in the discrete choice model for the chosen alternative.

With the disturbance terms, ω_i, having a multivariate logistic distribution then Hay (1980) demonstrates that:

$$E(\omega_{ij} | \omega_{j\ell} < \tau_{j\ell}) = \left[\frac{N-1}{N} \log P_j + \sum_{\substack{i=1 \\ i \neq j}} \frac{\log(P_i)}{-N} \left(\frac{P_i}{1 - P_i} \right) \right] (-1)^{N+1} \qquad (7.20)$$

so that the continuous choice model with H-D+M selectivity correction may be written:

$$E_j = X_j\beta + \sum_{\substack{i=j \\ i\neq j}}^{N} \frac{6}{\pi^2} \rho_{\omega_i u_j}\sigma_{u_j u_j}\left[\frac{N-1}{N}\log P_j + \frac{\log(P_i)}{N}\left(\frac{P_i}{1-P_i}\right)\right](-1)^{N+1} + v_j \quad (7.21)$$

The two stage estimation procedure involves estimating the discrete choice model and using the estimated probabilities in estimating (7.21). The correction formulas for the variance-covariance matrix associated with estimation of equation (7.21) are presented in Hay (1980) and Dubin and McFadden (1980).

2.3 SUMMARY

The essential difference between the H-L and H-D+M selectivity correction procedures is that the former is based on the construction of a bivariate distribution whereas, in the latter, the conditional expectations of u_j given the ϵ_i are assumed to be linear to begin with. A tractable full information likelihood function is available for the H-L method but not for the H-D+M method (Trost and Lee 1984). Moreover much of the appeal of the H-M+D method disappears when unranked data is used, since only one selectivity correction factor can be estimated. This consideration led to the H-L method being chosen for the work reported later in this chapter. Unreported models, however, using the H-D+M method did not substantially alter the conclusions to be drawn from the work. The similarity in empirical results to emerge from applying the two methods has also been noted by Hensher and Milthorpe (1985).

3. MIXTURE MODELS VERSUS SAMPLE SELECTIVITY MODELS

Pourier and Rudd (1981) have recently challenged the use of sample selectivity models in some contexts. The authors argue that an alternative model specification, dubbed the 'mixture model', may suffice for some analyses and offer advantages in terms of data parsimony. It is the purpose of this section to compare the mixture and sample selectivity models, demonstrate two types of inferences that can be drawn from sample selectivity models, and argue for the use of sample selectivity, rather than mixture models in estimating equation system (7.1) and (7.2). For ease of exposition most of this discussion refers to a binary choice context.

The mixture model can be summarily be specified as (Maddala 1983):

$$w_q = \tau(\alpha, Z_q) + \epsilon_q^* \quad (7.22)$$

$$y_q^S = X_q^S \beta^S + \bar{u}_q^S \quad (7.22)$$

$$y_q^N = X_q^N \beta^N + \bar{u}_q^N \quad (7.22)$$

where terminology is as set out in Section 2.1 but with the superscript S referring to the portion of the population from which the sample is drawn, the superscript N to the portion of the population excluded from the sample, and with the error term ϵ_q^* having a well-defined distribution over the entire population, $\bar{u}_q^S$ defined only for the subpopulation for which $\tau(\alpha, Z_q) \geq \epsilon_q^*$ ($I_q = 1$), and $\bar{u}_q^N$ defined only for the subpopulation for which $\tau(\alpha, Z_q) < \epsilon_q^*$ ($I_q = 0$).

The selection model can be written as:

$$w_q = \tau(\alpha, Z_q) + \epsilon_q^* \quad (7.23)$$

$$y_q^S = X_q^S \beta^S + u_q^S \quad (7.23)$$

$$y_q^N = X_q^N \beta^N + u_q^N \quad (7.23)$$

where ϵ_q^*, u_q^S and u_q^N have well defined distributions over the entire population.

In the selection model equation (7.23) is estimated in conditional form as:

$$E(y_q^S | I_q = 1) = X_q^S \beta^S + \sigma_{u_q^S \epsilon_q^*} E(\epsilon_q^* | \epsilon_q^* > -\tau(\alpha, Z_q)) \tag{7.23}$$

This permits two kinds of inferences to be made.

The first type of inferences concern the conditional distribution. To obtain such inferences assume that the variable of interest is contained in both X_q^S and Z_q and, for each vector, lies in the kth position. Then the expected effect of a slight change in the value of this variable on the conditional distribution is:

$$\frac{\partial E(y_q^S | I_q = 1)}{\partial x^S} = \beta_k + \frac{\partial \left[\sigma_{u_q^S \epsilon_q^*} E(\epsilon_q^* | \epsilon_q^* > -\tau(\alpha, Z_q))\right]}{\partial x_q^S} \tag{7.24}$$

Equation (7.24) provides an estimate of the effect of a change in x_{kq}^S on y_q^S, given that q lies in the sample subset S. Referring to the binary store choice/shopping expenditure example developed in the previous section, suppose that the variable of interest is 'store prices' and that this affects both shopping expenditure and the probability that store 1 (= S in equations (7.22) - (7.24)) will be chosen. Equation (7.24) provides an estimate of the effect of store prices on shopping expenditure given that store 1 has been chosen.

Inferences relating to the marginal distribution can be obtained from equation (7.23) as:

$$\frac{\partial E(y_q^S)}{\partial x_{kq}^S} = \beta_k \tag{7.25}$$

Again using the binary store choice/shopping expenditure example, equation (7.25) provides an estimate of the effect of store prices on shopping expenditure for an individual drawn at random from the population. This second estimate in essence removes the effect of store choice on shopping expenditure, that is, it is unconditional on store choice. The importance of this type of estimate is that β_1^S, estimated solely with data from store 1, can be used to provide an unbiased estimate of shopping expenditure for an individual drawn at random from the entire population.

It can be seen from the above that two types of inferences can be drawn from a sample selectivity model; one concerning the conditional distribution, the other concerning the marginal distribution. In contrast, because of the definition of $\bar{u}_q^S$, only conditional inferences can be made from the mixture model. The question, however, arises; 'In cases where only conditional inferences are required, is not the mixture model to be preferred for reasons of data parsimony?' This is the crux of Pourier and Rudd's argument. Duan et al. (1983) in their analysis of the effect of alternative health care insurance schemes on medical expenditures argue that theirs is one such case. It also seems likely that for analyses involving the use of unranked data, where a distinction is made only between chosen and non-chosen options in the discrete choice model, generally only conditional inferences will be required. The store choice model empirically developed in the latter part of this chapter utilises unranked data as do analyses conducted by Hensher and Milthorpe (1985) and Mannering (1985) among others. It is argued below, however, that even when only conditional inferences are required, there exist good reasons in most situations to favour the sample selectivity model.

The first reason refers to the generality of the two model structures. Statistically, given typical distributional assumptions concerning the error terms - for example, that they all normally distributed - the mixture model may be regarded as nested within the sample selectivity model with the testable parameter restrictions

$\sigma_{u_q^S \epsilon_q^*} = \sigma_{u_q^N \epsilon_q^*} = 0$. It should further be noted in this context, that for the common linear specification for the **unconditional** equation, an important difference between the mixture and sample selection models is that the latter results in a non-linear specification for the **conditional** equation. These points can be made more forcefully provided the Mardia-Lee transformation method to normality is accepted (Lee 1983), as the error term ϵ_q^* can then take a different distributional form to the u_q^k, k = S, N. On the other hand if the u_q^k are assumed to be non-normally and differently distributed, so that no nested relationship exists between the mixture model and sample selectivity model, then the preceding arguments hold no weight.

In the same vein more general structural models can be created that are consistent with the sample selectivity model than with the mixture model. For example, taking variants of the conditional indirect utility function used by Dubin and McFadden (1984 equation (10)) when analysing choice of water and space heating portfolios within the home, a mixture model version may be expressed as:

$$V_i = (\alpha_0 + \frac{\alpha_1}{\alpha_2} + \alpha_1 p_1 + \alpha_3 p_2 + \alpha_2 (Y - r_i) + \bar{u}_{1i}) \exp(\alpha_2 p_1) - \alpha_4 \log p_2 + \epsilon_i \qquad (7.26)$$

where p_1 and p_2 are the unit prices of electricity and gas, Y is income, r_i is the rental price of portfolio i, α_0, α_1, α_2, ..., α_4 are unknown parameters, the $\bar{u}_{1i}$ are error terms with zero conditional means and distributions defined for the subpopulation of portfolio consumers and the ϵ_i are independently and identically distributed error terms with unconditional means equal to zero, and variances equal to $(\sigma_{\epsilon_i \epsilon_i})^2$. After application of Roy's identity the electricity demand equation is:

$$C_i = \alpha_0 + \alpha_1 p_1 + \alpha_3 p_2 + \alpha_2 (Y - r_i) + \bar{u}_{1i} \qquad (7.27)$$

A sample selectivity version may be expressed as:

$$V_i = (\alpha_0 + \frac{\alpha_1}{\alpha_2} + \alpha_1 p_1 + \alpha_3 p_2 + \alpha_2 (Y - r_i) + u + \bar{u}_{1i}) \exp(\alpha_2 p_1) - \alpha_4 \log p_2 + \epsilon_i \quad (7.28)$$

where u is an extra error term defined for the entire population. The indirect utility function of equation (7.28) leads to an electricity demand equation:

$$C_i = \alpha_0 + \alpha_1 p_1 + \alpha_3 p_2 + \alpha_2 (Y - r_i) + u + \bar{u}_{1i} \quad (7.29)$$

When $\bar{u}_{1i} = 0$ and the ϵ_i and u are normally distributed, the discrete choice model (selection equation) can be estimated by multinomial probit and equation (7.29) by a variant of the methods outlined in Section 2.2 (see Terza 1985). This is the only model of the structural set (7.26) - (7.27) and (7.28) - (7.29) which is readily estimable.

Secondly, the question arises whether in the mixture model there exists any joint distribution for $(\bar{u}^k_q, \epsilon^*_q)$ that allows $\bar{u}^k_q$ and ϵ^*_q to be stochastically dependent or whether independence needs to be imposed. From Duan et al. (1984) it is possible to construct distributional forms that do result in the mixture model and allow $\bar{u}^k_q$ and ϵ^*_q to be stochastically dependent. The set of distributions satisfying these conditions, however, can be demonstrated as rather restrictive (Hay and Olsen 1984).

To summarise, the sample selection model must be used when inferences from the marginal distribution are required under conditions when only the conditional distribution is observed. The sample selection model should also be favoured, for reasons enunciated above, when the u^k_q are likely to be normally distributed or the u^k_q and ϵ^*_q are thought to have the same distributional form. However, when the distribution of the u^k_q is likely to be non-normal and of a different functional form to the ϵ^*_q and when inferences are only required from the conditional distribution, little guidance can be provided on the most appropriate choice of model.

Finally, it needs to be recognised that there exists an almost infinite potential for sample selectivity (Berk 1983). Even if a completely random sample is drawn from a defined population, that population will almost certainly represent some non-random subset of a more general population. To account for all selectivity, therefore, would lead to a sumptuous estimation process. On the other hand any universal application of uncorrected results obtained from a sample, may be misleading due to the presence of bias. The critical question is whether the bias is small enough to be safely ignored. The key to this question lies in the magnitude of the correlation between the error terms ϵ_q^* and u_q. If it can safely be assumed that this correlation is very small, then expediency may dictate that the sample selectivity term be ignored.

4. ESTIMATION OF AN INTEGRATED STORE CHOICE/SHOPPING EXPENDITURE MODEL

In the case of the shopping expenditure equations there is good reason to believe that the error terms in the discrete store choice (selection) model and the continuous shopping expenditure model may be correlated as they both stem from the same source, namely, analyst uncertainty about the indirect utility function. Also it seems reasonable to suppose that the error terms associated with the continuous shopping expenditure model will be normally distributed. The sample selectivity model is therefore favoured as an estimation method.

Actual estimation was done on a modified version of the model system specified in equations (7.1) and (7.2). In initial testing, the conditional indirect utility functions of equation (7.1) proved to be unstable with the data set available. Consequently a simplification was imposed and the conditional indirect utility functions respecified as:

$$V_{iq} = \Big[\alpha_3 - \alpha_2 \log (p_{iq}/\psi_{iq}) + \alpha_1 (Y_q - c_{iq}) + \alpha_4 (T - t_{iq}) \Big] (p_{iq})^{-\alpha_5} + \epsilon_{iq} \qquad (7.30)$$

which still yields the demand function shown in equation (7.2). In terms of economic theory equation (7.1) implies full use of the Fisher/Shell simple repackaging hypothesis. It may be regarded, however, that the inclusion of quality terms in equation (7.30), follows the more pragmatic course of expansion of the utility expression constant term, say,

$$\alpha_{i3}^{*} = \alpha_3 + \alpha_2 \sum_k \gamma_k \, b_{ikq}$$

with,

$$V_{iq} = \Big[\alpha_{i3}^{*} - \alpha_2 \log p_{iq} + \alpha_1 (Y_q - c_{iq}) + \alpha_4 (T - t_{iq}) \Big] (p_{iq})^{-\alpha_5} + \epsilon_{iq}$$

In the remainder of this chapter properties associated with equation system (7.30) and (7.2) are further explored. In Section 4.1 the exact form of model to be estimated is derived and the conditions discussed for equation (7.30) to represent a valid indirect utility function. Results from estimation are presented in Section 4.2.

4.1 TOWARDS AN ESTIMABLE MODEL AND THE DIEWERT CONDITIONS

In obtaining the estimated store choice model, the quality index of equation (4.6) and the perceived price transformation of equation (4.7) were inserted into the conditional indirect utility function of equation (7.30) which gave:

$$V_{iq} = \Big[\alpha_3 \, a_1^{\alpha_5} - \alpha_2 \, a_1^{\alpha_5} \log a_1 - \alpha_2 \, a_1^{\alpha_5} \, a_2 \log p^*_{iq}$$

$$+ \alpha_2 \, \gamma_1 \, a_1^{\alpha_5} \, SEL_{iq} + \alpha_2 \, \gamma_2 \, a_1^{\alpha_5} \, CONV_{iq}$$

$$+ \alpha_1 \, a_1^{\alpha_5} \, (Y_q - c_{iq}) + \alpha_4 \, a_1^{\alpha_5} \, (T - t_{iq}) \Big]$$

$$\times \, (p^*_{iq})^{a_2 \alpha_5} + \epsilon_{iq} \qquad (7.31a)$$

or,

$$V_{iq} = \Big[PAR(1) + PAR(2) \log p^*_{iq} + PAR(3) \, SEL_{iq}$$

$$+ PAR(4) \, CONV_{iq} + PAR(5) \, (Y_q - c_{iq})$$

$$+ PAR(6) \, (T - t_{iq}) \Big] \, p^{*\,PAR(7)}_{iq} + \epsilon_{iq} \qquad (7.31b)$$

where $PAR(1) = \alpha_3 \, a_1^{\alpha_5} - \alpha_2 \, a_1^{\alpha_5} \log a_1$, $PAR(2) = - \alpha_2 \, a_1^{\alpha_5} \, a_2$, $PAR(3) = \alpha_2 \, \gamma_1 \, a_1^{\alpha_5}$, $PAR(4) = \alpha_2 \, \gamma_2 \, a_1^{\alpha_5}$, $PAR(5) = \alpha_1 \, a_1^{\alpha_5}$, $PAR(6) = \alpha_4 \, a_1^{\alpha_5}$, and

$PAR(7) = a_2 \, \alpha_5$.

Where the ϵ_{iq} are assumed to be independently and identically distributed extreme value type 1 the store choice model takes the non-linear MNL form:

$$\text{Prob}\left\{I_q=j\right\} = \exp\left\{\left[\frac{PAR(1)}{\mu} + \frac{PAR(2)}{\mu}\log p^*_{jq} + \frac{PAR(3)}{\mu}SEL_{jq} + \frac{PAR(4)}{\mu}CONV_{jq} + \frac{PAR(5)}{\mu}(Y_q - c_{jq}) + \frac{PAR(6)}{\mu}(T - t_{jq})\right] p^{*\ PAR(7)}_{jq}\right\} \Big/ \sum_i^{N_q} \exp\left\{\left[\frac{PAR(1)}{\mu} + \frac{PAR(2)}{\mu}\log p^*_{iq} + \frac{PAR(3)}{\mu}SEL_{iq} + \frac{PAR(4)}{\mu}CONV_{iq} + \frac{PAR(5)}{\mu}(Y_q - c_{iq}) + \frac{PAR(6)}{\mu}(T - t_{iq})\right] \times p^{*\ PAR(7)}_{iq}\right\} \quad (7.32)$$

where $\mu = \sqrt{3}\left(\sigma_{\epsilon_i \epsilon_i}\right)/ 3.1416$ is the logistic positive scale factor.

Similarly the expanded form of the estimated continuous choice model is:

$$E_{qj} = \frac{\alpha_5\alpha_3}{\alpha_1} + \frac{\alpha_2}{\alpha_1} - \frac{\alpha_2\alpha_5}{\alpha_1}\log a_1 - \frac{\alpha_2\alpha_5}{\alpha_1}a_2\log p^*_{jq} + \frac{\alpha_2\alpha_5}{\alpha_1}\gamma_1 SEL_{jq} + \frac{\alpha_2\alpha_5}{\alpha_1}\gamma_2 CONV_{jq} + \alpha_5(Y_q - c_{jq}) + \frac{\alpha_4\alpha_5}{\alpha_1}(T - t_{jq}) - \frac{\sigma_{\eta^*_j u_j}}{(\sigma_{\eta^*_j \eta_j})^2} \times (\phi\,[\Phi^{-1}(\text{Prob}\{I_q = j\})])/(\text{Prob}\{I_q = j\}) + v_{qj} \quad (7.33a)$$

or,

$$E_{qj} = PAR(8) + PAR(9)\log p^*_{jq} + PAR(10)SEL_{jq} + PAR(11)CONV_{jq} + PAR(12)\,(Y_q - c_{jq}) + PAR(13)\,(T - t_{jq}) - PAR(14)\frac{\phi\,[\Phi^{-1}(\text{Prob}\{I_q = j\})]}{\text{Prob}\{I_q = j\}} + v_{qj} \quad (7.33b)$$

where $PAR(8) = \frac{\alpha_5\alpha_3}{\alpha_1} + \frac{\alpha_2}{\alpha_1} - \frac{\alpha_2\alpha_5}{\alpha_1} \log a_1$, $PAR(9) = -\frac{\alpha_2\alpha_5}{\alpha_1} a_2$, $PAR(10) = \frac{\alpha_2\alpha_5}{\alpha_1} \gamma_1$, $PAR(11) = \frac{\alpha_2\alpha_5}{\alpha_1} \gamma_2$, $PAR(12) = \alpha_5$, $PAR(13) = \frac{\alpha_4\alpha_5}{\alpha_1}$, $PAR(14) = \frac{\sigma_{\eta_j^* u_j}}{(\sigma_{\eta_j^*\eta_j^*})^2}$ and v_{qj} are new residuals with zero conditional means and zero conditional covariances for different q.

In total equation system (7.32) - (7.33) provides 14 parameter estimates. Associated with these estimates are 11 structural parameters $\alpha_1, \ldots, \alpha_5, a_1, a_2, \gamma_1, \gamma_2, \mu$ and $\sigma_{\eta_j^* u_j}/(\sigma_{\eta_j^*\eta_j^*})^2$. As will be seen, however, not all structural parameters can be identified. Conversely, because some structural parameter combinations are estimated more than once, internal consistency checks can be developed for the model system.

Five internal consistency checks can be constructed. These are:

(i) $$\frac{PAR(3)}{PAR(4)} = \frac{\gamma_1}{\gamma_2} = \frac{PAR(10)}{PAR(11)}$$

(ii) $$\frac{PAR(2)}{PAR(3)} = \frac{a_2}{\gamma_1} = \frac{PAR(9)}{PAR(10)}$$

(iii) $$\frac{PAR(6)}{PAR(5)} = \frac{\alpha_4}{\alpha_1} = \frac{PAR(13)}{PAR(12)}$$

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

(v) $$\text{PAR}(8) - \frac{\text{PAR}(2)}{\text{PAR}(5)} \times \frac{\text{PAR}(12)}{\text{PAR}(7)}$$

$$= \frac{\alpha_5 \ \alpha_3}{\alpha_1} - \frac{\alpha_5 \ \alpha_2}{\alpha_1} \times \log a_1$$

$$= \text{PAR}(11) \times \frac{\text{PAR}(1)}{\text{PAR}(4)}$$

These relationships mean that the original set of 13 equations can be reduced to 8 equations. Solution values can be found for the structural parameters:

$$\alpha_5 = \text{PAR}(12),$$

$$a_2 = \frac{\text{PAR}(7)}{\text{PAR}(12)},$$

$$\gamma_1 = \frac{\text{PAR}(3)}{\text{PAR}(2)} \times \frac{\text{PAR}(7)}{\text{PAR}(12)}, \text{ and}$$

$$\gamma_2 = \frac{\text{PAR}(4)}{\text{PAR}(2)} \times \frac{\text{PAR}(7)}{\text{PAR}(12)}.$$

The remaining structural parameters can be expressed as combinations of the estimated parameters, PAR(1) - PAR(7) and PAR(12), and a_1 and α_1. The overall solution vector is:

$$
\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ a_1 \\ a_2 \\ \gamma_1 \\ \gamma_2 \\ \mu \end{bmatrix}
=
\begin{bmatrix}
\alpha_1 \\
\dfrac{PAR(2)}{PAR(5)} \times \dfrac{PAR(12)}{PAR(7)} \\
\dfrac{PAR(1)}{PAR(5)} \alpha_1 + \dfrac{PAR(2)}{PAR(5)} \times \dfrac{PAR(12)}{PAR(7)} \alpha_1 \log a_1 \\
\dfrac{PAR(6)}{PAR(5)} \alpha_1 \\
PAR(12) \\
a_1 \\
\dfrac{PAR(7)}{PAR(12)} \\
\dfrac{PAR(3)}{PAR(2)} \times \dfrac{PAR(7)}{PAR(12)} \\
\dfrac{PAR(4)}{PAR(2)} \times \dfrac{PAR(7)}{PAR(12)} \\
\dfrac{\alpha_1}{PAR(5)} \quad \alpha_1 a_1^{PAR(12)}
\end{bmatrix}
$$

For the purposes of research reported here, however, the inability to identify all structural parameters represents no great handicap*.

---

* In the remainder of this Chapter, for reasons of notational economy, the normalisation is invoked $(\sigma_{\epsilon_i \epsilon_i})^2$

As noted in Chapter 3 for the function of equation (7.30) to represent a valid conditional indirect utility function it must conform to a number of conditions. Diewert (1974) has shown that indirect utility functions possess the following properties:

(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.

These conditions are now specified for the conditional indirect utility function shown in equation (7.30).

The first two conditions can be quickly disposed. Clearly equation (7.30) is continuous in the positive domain of prices and income. Also in the derivation of equation (7.30) all prices and income were normalised by the price of the Hicksian commodity (see Chapter 3). As a result equation (7.30) is homogeneous of degree zero in prices and income.

The conditions that V(.) be non-increasing in prices implies $\partial V / \partial p \leq 0$. For equation (7.30) for price p_i:

$$\frac{\partial V_{iq}}{\partial p_i} = \Big[\alpha_3 - \alpha_2 \log p_i + \alpha_2 \gamma_1 SEL_{iq} + \alpha_2 \gamma_2 CONV_{iq} + \alpha_1 (Y_q - c_{iq}) + \alpha_4 (T - t_{iq}) \Big] \alpha_5 p_i^{(\alpha_5 - 1)} + \alpha_2 p_i^{(\alpha_5 - 1)}$$

Substituting $a_1 p_{iq}^{* a_2}$ for p_i (see equation 4.7) yields:

= 1 and the term $\sqrt{3} / 3.1416$, neglected.

$$\frac{\partial V_{iq}}{\partial p_i} = \Big[\alpha_3 - \alpha_2 \log a_1 - \alpha_2 a_2 \log p^*_{iq} + \alpha_2 \gamma_1 SEL_{iq}$$

$$+ \alpha_2 \gamma_2 CONV_{iq} + \alpha_1 (Y_q - c_{iq}) + \alpha_4 (T - t_{iq}) \Big]$$

$$x \quad \alpha_5 a_1^{(\alpha_5 - 1)} \; p^{*}_{iq}{}^{a_2(\alpha_5 - 1)}$$

$$+ \alpha_5 a_1^{(\alpha_5 - 1)} \; p^{*}_{iq}{}^{a_2(\alpha_5 - 1)} \qquad (7.34)$$

Evaluation of equation (7.34) requires knowledge of all structural parameters. A simpler method of testing the non-increasing price condition is to note, by straightforward application of the chain rule,

$$\frac{\partial V_{iq}}{\partial p^*_{iq}} = \frac{\partial p_i}{\partial p^*_{iq}} \frac{\partial V_{iq}}{\partial p_i} \qquad (7.35)$$

Given that perceived prices are a monotonically increasing function of real prices, implying $\partial p_i / \partial p^*_{iq} > 0$, then the condition that $\partial V_{iq} / \partial p_i \leq 0$ implies that $\partial V_{iq} / \partial p^*_{iq} \leq 0$ for all p^*_{iq}. Also $\partial V_{iq} / \partial p^*_{iq}$ can be obtained through differentiation of the reduced form model of equation (7.31b).

$$\frac{\partial V_{iq}}{\partial p^*_{iq}} = \Big[PAR(1) + PAR(2) \log p^*_{iq} + PAR(3)\ SEL_{iq}$$

$$+ PAR(4)\ CONV_{iq} + PAR(5)\ (Y_q - c_{iq})$$

$$+ PAR(6)\ (T - t_{iq}) \Big] PAR(7)\ p^{*\,(PAR(7) - 1)}_{iq}$$

$$- PAR(2)\ p^{*\,(PAR(7) - 1)}_{iq} \qquad (7.36)$$

In general the value of $\partial V_{iq} / \partial p^*_{iq}$ will depend on the values attached to the parameter vector and all variables.

The conditional indirect utility function should also be non-decreasing in income, that is, $\partial V_{iq} / \partial Y_q \geq 0$ for q = 1, 2, ..., Q. From equation (7.31b)

$$\frac{\partial V_{iq}}{\partial Y_q} = PAR(5)\ p_{iq}^{*\ PAR(7)} \qquad (7.37)$$

Provided PAR(5) is positive V_{iq} will be non-decreasing in income for q = 1, 2, ..., Q.

Finally the quasi-convexity condition is derived. For the conditional indirect utility functions $V(p_i, B_{iq}, T - t_{iq}, Y - c_{iq}, \epsilon_{iq})$ this condition implies that the s_{ii} element of the Slutsky matrix be non-positive (e.g. Hausman 1981). This element represents the second partial derivative of the conditional cost function with respect to p_i. From equation (7.30), by inversion, the cost function is:

$$\begin{aligned} Y_{iq} = {} & V_{iq}\ \alpha_1^{-1}\ p_i^{\alpha_5} - \alpha_1^{-1} \Big[\alpha_3 - \alpha_2 \log p_i \\ & + \alpha_2\ \gamma_1\ SEL_{iq} + \alpha_2\ \gamma_2\ CONV_{iq} - \alpha_1\ c_{iq} \\ & + \alpha_4\ (T - t_{iq}) \Big] - \epsilon_{iq}\ \alpha_1^{-1}\ p_i^{\alpha_5} \end{aligned} \qquad (7.38)$$

Partially differentiating equation (7.38) w.r.t. p_i gives the conditional Hicksian demand function:

$$\begin{aligned} \frac{\partial Y_{iq}}{\partial p_i} & = h_{iq}\ (p_i, B_{iq}, \epsilon_{iq}, V_{iq}) \\ & = V_{iq}\ \alpha_1^{-1}\ \alpha_5\ p_i^{(\alpha_5 - 1)} - \alpha_1^{-1}\ \alpha_2\ p_i^{-1} \\ & \quad - \alpha_5\ \alpha_1^{-1}\ p_i^{(\alpha_5 - 1)}\ \epsilon_{iq} \end{aligned} \qquad (7.39)$$

and partially differentiating the conditional Hicksian demand function of equation (7.39) w.r.t. p_i gives the s_{ii} element of the Slutsky matrix:

$$\frac{\partial h_{iq}}{\partial p_i} = s_{ii} = (\alpha_5 - 1)\ \alpha_1^{-1}\ \alpha_5\ p_i^{-2} \Big[\alpha_3 - \alpha_2 \log p_i + \alpha_2\ \gamma_1\ SEL_{iq} + \alpha_2\ \gamma_2\ CONV_{iq} + \alpha_1\ (Y_q - c_{iq}) + \alpha_4\ (T - t_{iq}) \Big] + \alpha_1^{-1}\ \alpha_2\ p_i^{-2} \qquad (7.40)$$

where $s_{ii} \leq 0$ for equation (7.30) to represent a valid conditional indirect utility function.

Evaluation of equation (7.40) requires full knowledge of the structural parameters. Again, however, a relationship can be established between the quasi-convex condition for real prices and an analogous condition involving perceived prices. To derive this condition, the conditional cost function can generally be described by:

$$Y_{iq} = Y_{iq}\ (p_i,\ B_{iq},\ c_{iq},\ T - t_{iq},\ \epsilon_{iq},\ V_{iq})$$

$$= Y_{iq}\ (p_i(p_{iq}^*),\ B_{iq},\ c_{iq},\ T - t_{iq},\ \epsilon_{iq},\ V_{iq}) \qquad (7.41)$$

Taking the partial differential of equation (7.41) w.r.t. p_{iq}^* gives:

$$\frac{\partial Y_{iq}}{\partial p_{iq}^*} = \frac{\partial Y_{iq}}{\partial p_i}\ \frac{\partial p_i}{\partial p_{iq}^*} \qquad (7.42)$$

To represent a valid cost function equation (7.41) must be non-decreasing in prices, that is, $\partial Y_{iq}/\partial p_i \geq 0$ (see, for example, Varian 1984). Also $\partial p_i/\partial p^*_{iq}$ is greater than or equal to zero. Thus the condition, using perceived prices, for equation (7.41) to represent a valid cost function is $\partial Y_{iq}/\partial p^*_{iq} \geq 0$.

The second partial derivative of equation (7.41) w.r.t. p^*_{iq} is:

$$\frac{\partial^2 Y_{iq}}{\partial p^*_{iq}\,\partial p^*_{iq}} = \frac{\partial}{\partial p^*_{iq}}\left(\frac{\partial Y_{iq}}{\partial p_i}\;\frac{\partial p_i}{\partial p^*_{iq}}\right)$$

$$= \frac{\partial^2 Y_{iq}}{\partial p_i \partial p_i}\left(\frac{\partial p_i}{\partial p^*_{iq}}\right)^2 + \frac{\partial Y_{iq}}{\partial p_i}\;\frac{\partial^2 p_i}{\partial p^*_{iq}\partial p^*_{iq}} \qquad (7.43)$$

Note that $\dfrac{\partial^2 Y_{iq}}{\partial p_i\ \partial p_i} \leq 0$ is the quasi- convexity condition for real prices. Also $\left(\dfrac{\partial p_i}{\partial p^*_{iq}}\right)^2$ must be zero or positive. Thus the first term on the RHS of equation (7.43) must be negative for V_{iq} to represent a valid conditional indirect utility function with Y_{iq} as the corresponding conditional expenditure function. Given that $\dfrac{\partial Y_{iq}}{\partial p_{iq}}$ is positive, provided $\dfrac{\partial^2 p_i}{\partial p^*_{iq}\partial p^*_{iq}}$ is negative $\dfrac{\partial^2 Y_{iq}}{\partial p^*_{iq}\partial p^*_{iq}}$ must also be negative for V_{iq} to meet the quasi-convexity condition. On the other hand if $\frac{\partial^2 p_i}{\partial p^*_{iq} \partial p^*_{iq}}$ is positive the correct sign of $\frac{\partial^2 Y_{iq}}{\partial p^*_{iq} \partial p^*_{iq}}$ for V_{iq} to meet the quasi-convexity condition is indeterminate.

Referring to equation (4.7), the second partial derivative of p_i w.r.t. p^*_i is:

$$\frac{\partial^2 p_i}{\partial p^*_{iq} \, \partial p^*_{iq}} = (a_2 - 1) \; a_2 \; a_1 \; p^{*\,(a_2 - 2)}_{iq}$$

If a_2 is negative, for p_i to be monotonically increasing in p^*_{iq}, a_1 must be negative and $\frac{\partial^2 p_i}{\partial p^*_{iq} \, \partial p^*_{iq}}$ will also be negative. On the other hand if a_2 is positive a_1 must also be positive. For $\frac{\partial^2 p_i}{\partial p^*_{iq} \partial p^*_{iq}}$ to be then negative a_2 must lie between 0 and 1. An estimate of a_2 can easily be obtained by dividing PAR(7) by PAR(12).

Compiling these facts, for $V_{iq} = V_{iq}$ $(p_i(p^*_{iq}), B_{iq}, T - t_{iq}, Y_q - c_{iq}, \epsilon_i)$ to represent a valid conditional indirect utility function with Y_{iq} as the corresponding conditional expenditure function and $p_i = a_1 p^{*\,a_2}_{iq}$, then (from equation (7.43)) $\frac{\partial^2 Y_{iq}}{\partial p_i \, \partial p_i} \left(\frac{\partial p_i}{\partial p^*_{iq}}\right)^2$ will be negative. Further, $\frac{\partial Y_{iq}}{\partial p_i}$ will be positive. Provided p_i is a monotonically increasing function of p^*_{iq}, if $a_2 \leq 1$, $\frac{\partial^2 p_i}{\partial p^*_{iq} \, \partial p^*_{iq}}$

will be negative, $\dfrac{\partial Y_{iq}}{\partial p_i}$ $\dfrac{\partial^2 p_i}{\partial p^*_{iq}\,\partial p^*_{iq}}$ will be negative, and $\dfrac{\partial^2 Y_{iq}}{\partial p^*_{iq}\,\partial p^*_{iq}}$ must be negative. Conversely if $a_2 > 1$ then $\dfrac{\partial^2 p_i}{\partial p^*_{iq}\,\partial p^*_{iq}}$ will be positive, $\dfrac{\partial Y_{iq}}{\partial p_i}$ $\dfrac{\partial^2 p_i}{\partial p^*_{iq}\,\partial p^*_{iq}}$ will be positive and the sign of $\dfrac{\partial^2 Y_{iq}}{\partial p^*_{iq}\,\partial p^*_{iq}}$, indeterminant. That is, the quasi-convexity condition can be tested using perceived prices provided the estimate of $a_2 \leq 1$. Note, however, this is a weakened quasi-convexity test. From equation (7.43) with

$$\frac{\partial Y_{iq}}{\partial p_i} \quad \frac{\partial^2 p_i}{\partial p^*_{iq}\,\partial p^*_{iq}} < \frac{\partial^2 Y_{iq}}{\partial p_i \;\partial p_i} \quad \left(\frac{\partial p_i}{\partial p^*_{iq}}\right)^2$$

it is possible for $\dfrac{\partial^2 Y_{iq}}{\partial p^*_{iq}\,\partial p^*_{iq}}$ to be negative, but for V_{iq} (.) not to be quasi-convex in real prices. That $\dfrac{\partial^2 Y_{iq}}{\partial p^*_{iq}\,\partial p^*_{iq}}$ be negative when $a_2 \leq 1$ is a necessary, but not sufficient, condition for V_{iq} (.) to be quasi-convex in real prices.

The conditional expenditure functions corresponding to the reduced form conditional indirect utility functions are

$$\begin{aligned} Y_{iq} = {} & PAR(5)^{-1}\, p^{*\;-\,PAR(7)}_{iq}\, V_{iq} - PAR(5)^{-1} \Big[PAR(1) \\ & + PAR(2) \log p^*_{iq} + PAR(3)\, SEL_{iq} + PAR(4)\, CONV_{iq} \\ & - PAR(5)\, c_{iq} + PAR(6)\, (T - t_{iq}) \Big] \\ & - \epsilon_{iq}\, PAR(5)^{-1}\, p^{*\;-\,PAR(7)}_{iq} \end{aligned}$$

with partial second derivatives w.r.t. p^*_{iq} given by:

$$\frac{\partial Y_{iq}}{\partial p^*_{iq} \, \partial p^*_{iq}} = (PAR(7) + 1) \frac{PAR(7)}{PAR(5)} p^{*\,(-2)}_{iq} \Big[PAR(1) + PAR(2) \log p^*_{iq} + PAR(3)\, SEL_{iq} + PAR(4)\, CONV_{iq} + PAR(5)\, (Y_q - c_{iq}) + PAR(6)\, (T - t_{iq}) \Big] + \frac{PAR(2)}{PAR(5)} p^{*\,(-2)}_{iq} \qquad (7.44)$$

4.2 ESTIMATION RESULTS

Data for estimating the model system were 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 the preceding trip originating from home or the subsequent trip ending at the home. 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 sifting was done on the basis of items of socio-economic information provided.

From this process two data sets, used for estimation, were compiled. One set only included, for each household, the maximum expenditure store. The other set used all the stores remaining after editing. For the second data set, in effect, each store choice by a household was treated independently. The number of households remaining after editing were 102. The second data set included 236 store choices. In both cases the choice set for any 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.

Discrete choice model parameter estimates for the 'maximum expenditure' and 'all stores' data sets are shown in Tables 7.1 and 7.2, respectively. The choice set comprised the store visited (i.e. the chosen store) plus two randomly selected stores from the set of stores provided. Estimation was achieved by a modified version of the software package 'TROMP' (Sparmann and Daganzo 1982). The modifications, effected by the author, principally involved enabling the program to handle explicitly varying choice set sizes and expansion of the number of attributes that could be included in the analysis.

The selectivity correction terms in the continuous shopping expenditure models (see Tables 7.3 and 7.4) were obtained by applying the H-L method described in Section 2.1. A listing of the computer program used to calculate these terms and the corrected standard errors is contained in Appendix 7B. The significance of the selectivity term for the 'maximum expenditure' data set confirms a priori suspicions concerning the possibility that the error terms in the store choice and shopping expenditure models would be correlated. It is also interesting that the selectivity correction factor is significant in the relatively ill specified model system. From model statistics the unobserved influences are playing a larger role in the 'maximum expenditure' data set. Thus the possibilities for correlation are enlarged.

A number of factors serve to engender confidence in the models. Firstly, virtually all variables took on their anticipated signs. The sole exception to this was the estimate for PAR(10) obtained from the maximum expenditure stores data set. It should be noted that this parameter estimate, as with other expenditure model parameter estimates using the maximum expenditure stores data set, is

TABLE 7.1

ESTIMATED STORE CHOICE MODEL : MAXIMUM EXPENDITURE STORE DATA SET

Form of conditional indirect utility functions*:

$$\overline{V}_{iq} = \Big[PAR(1) + PAR(2) \log (GPRICE_{dq}) + PAR(3)\ GSEL_{dq} + PAR(4)\ GCONV_{dq} + PAR(5)\ (HINCOME_{q} - TCOST_{mdq}) + PAR(6)\ (60-TTIME_{mdq}) \Big] GPRICE_{dq}^{\ PAR(7)}$$

Parameter	Parameter Estimate	Standard Error	T - Statistic
PAR(1)	37.4923	6.4866	5.78
PAR(2)	25.0037	6.0689	4.12
PAR(3)	1.8270	0.3980	4.59
PAR(4)	2.2422	0.3392	6.61
PAR(5)	0.0004	0.0022	0.18
PAR(6)	0.0814	0.0252	3.23
PAR(7)	-0.5356	0.0226	-23.75

ρ^2	0.065
% correctly predicted	
- at zero	29
- at convergence	39

*Note: Variables defined in Table 4.9.

TABLE 7.2

ESTIMATED STORE CHOICE MODEL - ALL STORES DATA SET

Form of conditional indirect utility functions*:

$$\overline{V}_{iq} = \Big[PAR(1) + PAR(2) \log (GPRICE_{dq}) + PAR(3)\ GSEL_{dq} + PAR(4)\ GONV_{dq} + PAR(5)\ (HINCOME_q - TCOST_{mdq}) + PAR(6)\ (60\text{-}TTIME)_{mdq} \Big] GPRICE_{dq}^{\ PAR(7)}$$

Parameter	Parameter Estimate	Standard Error	T - Statistic
PAR(1)	6.3813	0.3578	17.84
PAR(2)	-0.9051	0.3428	-2.64
PAR(3)	0.0344	0.0430	0.80
PAR(4)	0.1418	0.1028	1.38
PAR(5)	0.0059	0.0021	2.79
PAR(6)	0.0901	0.0213	4.23
PAR(7)	-0.0301	0.0032	-9.55

ρ^2	0.152
% correctly predicted:	
at zero	29
at convergence	45

*Note: Variables defined in Table 4.9.

TABLE 7.3

ESTIMATED SHOPPING EXPENDITURE MODEL - MAXIMUM EXPENDITURE STORE DATA SET

Form of demand equation*:

$$EXPEND_{iq} = PAR(8) + PAR(9) \log (GPRICE_{dq}) + PAR(10)\ GSEL_{dq}$$

$$+ PAR(11)\ GCONV_{dq} + PAR(12)\ (HINCOME_{q} - TCOST_{mdq})$$

$$+ PAR(13)\ (60-TTIME_{mdq}) - PAR(14)\ SCLEE_{mdq}$$

Parameter	Parameter Estimate	Uncorrected Standard Error	Uncorrected T-Statistic	Corrected T-Statistic
PAR(8)	13.3567	15.0041	0.89	0.80
PAR(9)	-10.8048	7.0291	-1.54	-1.41
PAR(10)	-0.6979	2.1496	-0.32	-0.29
PAR(11)	4.0675	2.1035	1.93	1.79
PAR(12)	0.0013	0.0054	0.25	0.22
PAR(13)	0.1029	0.2348	0.44	0.39
PAR(14)	5.0011	2.1738	2.30	2.01

$R^2 = 0.073$

*Note: Variables defined in Table 4.9.

TABLE 7.4

ESTIMATED SHOPPING EXPENDITURE MODEL - ALL STORES DATA SET

Form of demand equation*:

$$EXPEND_{iq} = PAR(8) + PAR(9)\ \log\ (GPRICE_{dq}) + PAR(10)\ GSEL_{dq}$$

$$+ PAR(11)\ GCONV_{dq} + PAR(12)\ (HINCOME_{q} - TCOST_{mdq})$$

$$+ PAR(13)\ (60-TTIME_{mdq}) - PAR(14)\ SCLEE_{mdq}$$

Parameter	Parameter Estimate	Uncorrected Standard Error	Uncorrected T-Statistic	Corrected T-Statistic
PAR(8)	3.9858	0.8134	4.90	4.88
PAR(9)	-10.3139	2.8388	3.63	3.62
PAR(10)	2.1661	1.0185	2.13	2.12
PAR(11)	3.6611	0.8709	4.20	4.19
PAR(12)	0.0058	0.2254	2.58	2.57
PAR(13)	0.1210	0.1677	0.72	0.72
PAR(14)	1.2481	1.6408	0.76	0.74

$R^2 = 0.217$

*Note: Variables defined in Table 4.9.

insignificantly different from zero at normal confidence levels. The insignificance of expenditure model parameter estimates using this data set can be attributed to the small sample size and reduced variation in observed expenditures caused by taking only the maximum expenditure store for each household.

Secondly, both sets of models passed the internal consistency tests derived in Section 4.1. Results from these tests are presented in Table 7.5. In all cases the internal consistency checks of parameter estimates fall within 95% confidence intervals *.

Thirdly, the estimated conditional indirect utility functions meet most of the Diewert conditions. In particular:

(i) The non-decreasing income conditions is met in all the estimated conditional indirect utility functions for both data sets, since the estimates attached to PAR(5) are positive.

(ii) The non-increasing price condition was met for all conditional indirect utility functions for every data point in both samples. Evaluated at sample means for the maximum stores data set
$\partial V_j / \partial p_j^* = -3.5482$ for chosen stores and
$\partial V_i / \partial p_i^* = -3.1268$ for non-chosen stores, while for the 'all stores' data set the corresponding figures are -0.3305 and -0.22889, respectively.

---

* Unfortunately these confidence intervals are often fairly wide due to all tests involving the use of ratios. The variance associated with A/B where A and B are random variables is:

$$\text{var}\ (A/B) = \frac{1}{B^4}\left\{ B^2\ \text{var}\ (A) + A^2\ \text{var}\ (B) - 2AB\ \text{COV}\ (AB)\right\}$$

For the calculations presented in Table 6 where the parameter estimates were from different models (e.g. PAR(11) and PAR(4) in test 4) the covariance term was assumed equal to zero.

TABLE 7.5

INTERNAL CONSISTENCY CHECKS OF THE INTEGRATED STORE CHOICE/SHOPPING EXPENDITURE MODEL SYSTEMS

TEST DESCRIPTION	TEST RESULTS			
	MAXIMUM EXPENDITURE STORES DATA SET		ALL STORES DATA SET	
	LHS point value (95% confidence interval)	RHS point value (95% confidence interval)	LHS point value (95% confidence interval)	RHS point value (95% confidence interval)
$\frac{PAR(3)}{PAR(4)} = \frac{PAR(10)}{PAR(11)}$	0.8148 (0.3552 to 1.2744)	-0.1716 (-1.2690 to 0.9258)	0.2426 (-0.5112 to 0.9964)	0.5917 (-0.1194 to 1.3027)
$\frac{PAR(2)}{PAR(3)} = \frac{PAR(9)}{PAR(10)}$	13.6856 (3.5924 to 23.7788)	15.4819 (76.1049 to 107.0687)	-26.3110 (-101.8792 to 49.2572)	-4.7615 (-9.1098 to -0.4132)
$\frac{PAR(6)}{PAR(5)} = \frac{PAR(13)}{PAR(12)}$	180.8889 (-2774 to 3136)	79.1539 (-637.113 to 795.4205)	15.2746 (1.9660 to 28.5832)	20.8621 (-1569 to 1610)
$\frac{PAR(11)}{PAR(4)} = \frac{PAR(12)}{PAR(5)}$	1.8141 (-0.1018 to 3.7300)	3.25000 (41.3407 to 47.8407)	25.8188 (-12.7924 to 64.4299)	0.9831 (-0.0436 to 2.0098)
$PAR(8) - \frac{PAR(2)}{PAR(5)} \times \frac{PAR(12)}{PAR(7)} = PAR(11) \times \frac{PAR(1)}{PAR(4)}$	165.23 (-2181 to 2511	68.0184 (-594.46 to 730.49)	-25.5741 (-68.4193 to 17.2711)	164.7748 (-80.1860 to 409.7356)

(iii) The estimate of a_2 was -412.00 for the maximum expenditure stores data set and -5.0167 for the 'all stores' data set. Since these two estimates lie within the interval $-\infty < a_2 \leq 1$, the quasi-convexity condition can be tested on both data sets using perceived prices. For the 'all stores' data set the weakened quasi-convexity condition was met for all conditional indirect utility functions for every data point in the sample. Evaluated at sample means $\frac{\partial^2 Y_{iq}}{\partial p^*_{iq} \partial p^*_{iq}}$ = -14.6188 for chosen stores and -11.3508 for non-chosen stores. For the maximum expenditure stores' data set the quasi-convexity condition was met in less than 5% of the sample points. Evaluated at sample means $\frac{\partial^2 Y_{iq}}{\partial p^*_{iq} \partial p^*_{iq}} = 1908.25$ for chosen stores and 1579.88 for non-chosen stores. This is a disappointing result. Other authors who have tested indirect utility functions for this condition have also obtained indifferent results (e.g. Wales and Woodland 1977, Brownstone 1980). Brownstone (1980) following the methods of Lau (1978) discusses a constrained estimation procedure that ensures the quasi-convexity condition is met.

Elasticity estimates associated with the two model sets are displayed in Tables 7.6 and 7.7. For the discrete choice model the perceived price elasticity estimate satisfies:

TABLE 7.6

DISCRETE CHOICE MODEL DIRECT ELASTICITY ESTIMATES

Description	Elasticity estimate 'maximum expenditure stores' data set	Elasticity estimate 'all stores' data set
perceived price elasticity	-2.17	-0.75
perceived selection elasticity	0.59	0.04
perceived store convenience elasticity	0.88	0.23
travel cost elasticity	0.00	0.00
travel time elasticity	0.01	0.05

TABLE 7.7

SHOPPING EXPENDITURE ELASTICITY ESTIMATES

Description	Elasticity estimate 'maximum expenditure stores' data set	Elasticity estimate 'all stores' data set
Elasticities of shopping expenditure with a given store:		
with respect to perceived prices	-0.59	-0.85
with respect to perceived selection	-0.10	0.31
with respect to perceived store convenience	0.68	0.67
with respect to income	0.06	0.28
with respect to time availability	0.30	0.41
Elasticities of expected shopping expenditure with flexible store choices:		
with respect to perceived prices	-2.85	-1.49
with respect to perceived selection	0.58	0.35
with respect to perceived store convenience	0.84	0.71
with respect to income	0.06	0.32
with respect to time availability	0.39	0.60

$$\frac{\partial \log \text{Prob}\{I_q = i\}}{\partial \log p^*_{iq}} = \Big\{[\text{PAR}(1) + \text{PAR}(2) \log p^*_{iq} + \text{PAR}(3)\ \text{SEL}_{iq} + \text{PAR}(4)\ \text{CONV}_{iq} + \text{PAR}(5)\ (Y_q - c_{iq}) + \text{PAR}(6)\ (T - t_{iq})] \times \text{PAR}(7)\ p^{*\ \text{PAR}(7)}_{iq} + \text{PAR}(2) \times p^{*\ \text{PAR}(7)}_{iq}\Big\} (1 - \text{Prob}\{I_q = i\}) \tag{7.45}$$

Elasticity estimates for the other discrete choice model variables can be similarly derived. Discrete choice elasticity estimates are displayed in Table 7.6. All estimates, except for perceived prices for the 'maximum expenditure stores' data set, are indicative of inelastic demand. The changes in elasticity estimates between the two data sets are in the direction that ad hoc reasoning suggests. Specifically for the higher expenditure data set the store attributes become relatively more important, while for the lower expenditures data set sampled individuals place relatively more weight on travel characteristics.

Two sets of elasticity estimates are provided for the shopping expenditure models. In the spirit of Dubin and McFadden (1984) the first set of elasticity estimates are calculated to correspond to short-run responses conditional on a particular store choice *. Although strictly incorrect, in calculating these elasticities the selectivity term has been held constant, under the assumption that the representative individual is locked into his current store choice in the short run.

---

* The 'short run' referred to here is likely to be of considerably less duration than that used by Dubin and McFadden in their study of water-heat space-heat portfolio choice and electricity demand. It is, however, still possible to imagine that store choices are in the short run inflexible due to factors such as individual inertia, household operating constraints, knowledge constraints, etc.

The second set of elasticity estimates are based on flexible store choices. They are calculated by using expected store expenditures and taking the selectivity term into account. Specifically they satisfy

$$\frac{\partial \log E_{iq}}{\partial \log x_{ikq}} = \left[\beta_k - \sigma_{u_i \eta_i^*} \frac{\partial \left[\Phi^{-1} (\text{Prob}\{I_q = i\}) \right]}{\partial x_{ikq}} \right] \times \frac{\partial \text{Prob}\{I_q = i\}}{\partial x_{ikq}} \frac{x_{ikq}}{E_{ikq}} \qquad (7.44)$$

where x_{iq} is one of the independent variables with β_k as its associated parameter in the expenditure model.

Not surprisingly, grocery shopping expenditure is characterised by inelastic demands. The only elasticity estimate greater than one is for perceived prices with flexible store choices. Naturally the longer run elasticities all exceed the short run elasticities. The time elasticity can be interpreted as the change in shopping expenditure that could be expected given that more time was available in which to shop. It can be seen from the models and elasticity estimates that store attributes (and even travel time and travel cost) not only impact the choice of store but also shopping expenditure once a store has been chosen.

5. CONCLUSION

An important aspect of the theory developed in Chapter 3 was the fusing of shopping destination choices with the shopping expenditure decisions made by individuals. This chapter has involved estimating an integrated store choice/shopping expenditure model system, consistent with that theory.

A good portion of this chapter was devoted to an in depth investigation of characteristics of the estimated model system. Because this system was soundly based on a well specified economic theory of shopping behaviour, a number of tests could be applied to the models that are unavailable when ad hoc approaches are used. An encouraging aspect was that the estimated model systems conformed to most of the conditions derived from the theoretical framework.

The insight offered by the integrated system is considerably richer than that offered by past modelling efforts in this area. The interrelated nature of store choice/shopping expenditure decisions is clearly evident in the empirical estimates obtained from the model system. In fact, in one data set used for estimation, this interrelatedness took an extremely complex form, but one which could be competently treated using advanced statistical techniques.

The estimated models demonstrated the importance of store attributes and time and income constraints on shopping expenditure. In this they lie in sharp relief to previously estimated models of shopping expenditure which have principally been based on individual or household socio-economic characteristics. One potentially significant way to improve the models estimated in this Chapter would be to add some socio-economic variables. The astute reader will be able to discern how this might be achieved within the integrated modelling framework. An area of statistical improvement would be to use full information maximum likelihood to simultaneously estimate all parameters. This would lead to more efficient estimates. Further, this approach permits all structural parameters to be recovered, thus enabling, for instance, a translation of perceived prices into real prices.

APPENDIX 7A

THE HECKMAN-LEE TWO-STAGE SELECTIVITY CORRECTION METHOD: DERIVATION OF THE VARIANCE/COVARIANCE MATRIX FOR THE CONTINUOUS CHOICE MODEL

In this Appendix I derive the variance/covariance matrix for the regression model with selectivity correction shown in equation (7.14). The methods used to derive the variance/covariance matrix follow those of Amemiya (1978), also used by Lee et al. (1980) and described by Maddala (1983). These authors, however, derive the variance/covariance matrix for a linear-in-the-parameters binary probit model. In this appendix the variance/covariance matrix derived is for the generalised selectivity correction factor constructed from non-linear-in-the-parameters multinomial logit model.

Consider a multinomial logit model given by:

$$D(\alpha, Z_{jq}) = \frac{\exp[\overline{V}(\alpha, Z_{jq})]}{\sum_{i}^{N} \exp[\overline{V}(\alpha, Z_{iq})]} \qquad \text{(A7.1)}$$

where α is a vector of parameters ($\alpha = \alpha_1, \alpha_2, \ldots, \alpha_R$), Z_i is Q x M matrix of variables pertaining to alternative i (i = 1,2,....,j,....,N),

$$Z_i = \begin{bmatrix} z_{i11} & z_{i12} & \cdots & z_{i1M} \\ z_{i21} & z_{i22} & \cdots & z_{i2M} \\ \vdots & \vdots & & \vdots \\ z_{iQ1} & z_{iQ2} & \cdots & z_{iQM} \end{bmatrix}$$

$\overline{V}_{iq}$ are unspecified functions comprising parameters α and variables Z_{iq}, and a normalisation is imposed so that $\overline{V}(\alpha, Z_{NQ}) \equiv 1$. Also note that the use of a single generic parameter vector in equation (A7.1), rather than separate parameter vectors pertaining to each alternative, does not imply any restriction on the generality of the model since alternative-specific effects can be introduced by defining some of the Z_{im} to be zero on all except one alternative.

From section 2.1, define:

$$J(\alpha, Z_{jq}) = \Phi^{-1}[D(\alpha, Z_{jq})] = \Phi^{-1}\left(\frac{\exp[\overline{V}(\alpha, Z_{jq})]}{\Sigma \exp[\overline{V}(\alpha, Z_{iq})]}\right) \quad \text{(A7.2)}$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution function. Given store j is chosen, expected shopping expenditure is:

$$E_{jq} = X_{jq}\beta - \sigma_{u_j u_j} \rho_{\eta_j^* u_j} \phi[J(\alpha, Z_{jq})]/D(\alpha, Z_{jq}) + v_{jq} \quad \text{(A7.3)}$$

where E_{jq} is expenditure by individual q at store j, β is a vector of parameters ($\beta' = \beta_1, \beta_2, \ldots, \beta_P$), X_j is a Q_j x P matrix of variables,

$$X_j = \begin{bmatrix} x_{j11} & x_{j12} & \cdots & x_{j1P} \\ x_{j21} & x_{j22} & \cdots & x_{j2P} \\ \vdots & \vdots & & \vdots \\ x_{jQ_j1} & \cdots & & x_{jQ_jP} \end{bmatrix}$$

Q_j is the number of individuals choosing store j (i.e. for which $I_q = j$), $(\sigma_{u_j u_j})^2$ is the variance of u_j (see section 2.1), $\rho_{\eta_j^* u_j}$ is the correlation of η_j^* with u_j (see section 2.1), ϕ is the density function of the standard normal and v_{jq} is an error term with $E(v_{jq}|I_q = j) = 0$.

In estimating equation (A7.3) using the two stage procedure α is replaced by the estimated parameter vector,

$\hat{\alpha}$, obtained from the multinomial logit model so that:

$$E_{jq} = X_{jq}\beta - \sigma_{u_j u_j}\rho_{\eta_j^* u_j}\,\phi[J(\hat{\alpha},Z_{jq})]/D(\hat{\alpha},Z_{jq}) + \tilde{v}_{jq} \qquad \text{(A7.4)}$$

where:

$$\tilde{v}_{jq} = v_{jq} + \sigma_{u_j u_j}\rho_{\eta_j^*}\,u_j \left[\frac{\phi[J(\hat{\alpha},Z_{jq})]}{D(\hat{\alpha},Z_{jq})} - \frac{\phi[J(\alpha,Z_{jq})]}{D(\alpha,Z_{jq})}\right]$$

Taking a Taylor series expansion of $\dfrac{\phi[J(\hat{\alpha},Z_{jq})]}{D(\hat{\alpha},Z_{jq})}$

around α and subtracting $\dfrac{\phi[J(\alpha,Z_{jq})]}{D(\alpha,Z_{jq})}$ we obtain:

$$\frac{\phi[J(\hat{\alpha},Z_{jq})]}{D(\hat{\alpha},Z_{jq})} - \frac{\phi[J(\alpha,Z_{iq})]}{D(\alpha,Z_{jq})} = B_\alpha(\hat{\alpha} - \alpha)$$

where B_α is an Q_j x R matrix:

$$B_\alpha = \begin{bmatrix} \dfrac{\partial\{\phi[J(\alpha,Z_{j1})]/D(\alpha,Z_{j1})\}}{\partial\alpha_1} & \cdots\cdots & \dfrac{\partial\{\phi[J(\alpha,Z_{j1})]/D(\alpha,Z_{j1})\}}{\partial\alpha_R} \\ \vdots & & \vdots \\ \dfrac{\partial\{\phi[J(\alpha,Z_{jQ_j})]/D(\alpha,Z_{jQ_j})\}}{\partial\alpha_1} & \cdots\cdots & \dfrac{\partial\{\phi[J(\alpha,Z_{jQ_j})]/D(\alpha,Z_{jQ_j})\}}{\partial\alpha_R} \end{bmatrix}$$

Denoting,

$$b_{\alpha_{rjq}} = \frac{\partial\{\phi(J(\alpha,Z_{jq}))[D(\alpha,Z_{jq})]^{-1}\}}{\partial\alpha_r}$$

and expanding by application of the chain and product rules:

$$b_{\alpha_{rjq}} = \frac{\partial\{\phi[J(\alpha,Z_{jq})]\}}{\partial[J(\alpha,Z_{jq})]} \; \frac{\partial[J(\alpha,Z_{jq})]}{\partial\alpha_r} \; \{D(\alpha,Z_{jq})\}^{-1}$$

$$-\phi[J(\alpha,Z_{jq})]\;[D(\alpha,Z_{jq})]^{-2}\;\frac{\partial\{D(\alpha,Z_{jq})\}}{\partial\alpha_r} \qquad \text{(A7.5)}$$

Recalling that:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\;(-x^2/2)$$

$$\text{so} \quad \frac{\partial[\phi(x)]}{\partial x} = -x\;\phi(x) \qquad \text{(A7.6)}$$

Applying the result in (A7.6), equation (A7.5) can be simplified to:

$$b_{\alpha_{rjq}} = \frac{-\phi[J(\alpha,Z_{qj})]}{D(\alpha,Z_{jq})} \left\{ J(\alpha,Z_{qj})\;\frac{\partial[J(\alpha,Z_{qj})]}{\partial\alpha_r} \right.$$

$$\left. +[D(\alpha,Z_{jq})]^{-1}\;\frac{\partial\{D(\alpha,Z_{jq})\}}{\partial\alpha_r} \right\} \qquad \text{(A7.7)}$$

Examining the terms in (A7.7) it is a relatively simple matter to show that:

$$\frac{\partial D(\overline{V}_{jq})}{\partial\alpha_r} = D(\overline{V}_{jq}) \left\{ \frac{\partial\overline{V}_{jq}}{\partial\alpha_r} - \sum_k D(\overline{V}_{kq})\;\frac{\partial\overline{V}_{kq}}{\partial\alpha_r} \right\} \qquad \text{(A7.8)}$$

where $\overline{V}_{jq} = \overline{V}_{jq}(\alpha,Z_{jq})$.

Also a standard result for the inverse, $x=f^{-1}(y)$, of a function $y=f(x)$ is: $\frac{dx}{dy} = \frac{1}{df(x)/dx}$

It follows that:

$$\frac{\partial[J(\alpha,Z_{jq})]}{\partial\alpha_r} = \frac{\partial\{\Phi^{-1}[D(\alpha,Z_{jq})]\}}{\partial\{D(\alpha,Z_{jq})\}} \frac{\partial\{D(\alpha,Z_{jq})\}}{\partial\alpha_r}$$

$$= \left[\frac{\partial\{\Phi[J(\alpha,Z_{jq})]\}}{\partial\{J(\alpha,Z_{jq})\}}\right]^{-1} \frac{\partial\{D(\alpha,Z_{jq})\}}{\partial\alpha_r}$$

$$= \left[\phi[J(\alpha,Z_{jq})]\right]^{-1} \frac{\partial\{D(\alpha,Z_{jq})\}}{\partial\alpha_r} \qquad \text{(A7.9)}$$

By inserting (A7.8) and (A7.9) into (A7.7) then,

$$b_{\alpha_{rjq}} = -\left[J(\alpha,Z_{qj}) + \frac{\phi[J(\alpha,Z_{qj})]}{D(\alpha,Z_{qj})}\right]\left[\frac{\partial\bar{V}_j}{\partial\alpha_r} - \sum_i D(\bar{V}_i)\frac{\partial\bar{V}_i}{\partial\alpha_r}\right] \qquad \text{(A7.10)}$$

Define G_j as a Q_j x (P + 1) matrix, $G_j = (X_j, \frac{-\Phi[J(\alpha,Z_{jq})]}{D(\alpha,Z_{jq})})$,

or,

$$G_j = \begin{bmatrix} x_{j11} & x_{j12} & \cdots\cdots & x_{j1P} & \dfrac{-\Phi[J(\alpha,Z_{j1})]}{D(\alpha,Z_{j1})} \\ x_{j21} & x_{j22} & \cdots\cdots & x_{j2P} & \dfrac{-\Phi[J(\alpha,Z_{j2})]}{D(\alpha,Z_{j2})} \\ \vdots & \vdots & & \vdots & \\ x_{jQ_j1} & x_{jQ_j2} & \cdots\cdots & x_{jQ_jP} & \dfrac{-\Phi[J(\alpha,Z_{jQ_j})]}{D(\alpha,Z_{jQ_j})} \end{bmatrix}$$

It follows from standard asymptotic theory that:

$$\begin{bmatrix} \widehat{\beta}_j \\ \widehat{\sigma_{u_ju_j}\rho_{\eta_j^*u_j}} \end{bmatrix} - \begin{bmatrix} \beta_j \\ \sigma_{u_ju_j}\rho_{\eta_j^*u_j} \end{bmatrix}$$

$$\stackrel{A}{=} (G_j'G_j)^{-1}G_j'\left[v_j + \sigma_{u_ju_j}\rho_{\eta_j^*u_j}\; B_\alpha(\hat{\alpha} - \alpha)\right] \qquad \text{(A7.11)}$$

where the notation $\stackrel{A}{=}$ means that the two expressions have the same asymptotic distribution. The asymptotic covariance matrix is therefore:

$$\text{var}\begin{bmatrix} \widehat{\beta}_j \\ \widehat{\sigma_{u_ju_j}\rho_{\eta_j^*u_j}} \end{bmatrix}$$

$$= (G_j'G_j)^{-1}G_j'\Big[\text{var}(v_j) + (\sigma_{u_ju_j}\rho_{\eta_j^*u_j})^2 B_\alpha\ \text{var}\ (\hat{\alpha})\ B_\alpha'$$
$$+ \sigma_{u_ju_j}\rho_{\eta_j^*u_j}\ B_\alpha\ \text{cov}(\hat{\alpha},v_j') + \sigma_{u_ju_j}\rho_{\eta_j^*u_j}\ \text{cov}(\alpha',v_j)B_\alpha'\Big]$$
$$\times\ G_j(G_j'G_j)^{-1} \qquad \text{(7.12)}$$

Define C_j as a Q_j x Q_j diagonal matrix with the qth diagonal term as:

$$\frac{\Phi[J(\alpha,Z_{jq})]}{D(\alpha,Z_{jq})}\left(\frac{\Phi[J(\alpha,Z_{jq})]}{D(\alpha,Z_{jq})} - J(\alpha,Z_{jq})\right)$$

Then as Lee (1983), following the results of Johnson and Kotz (1972), has shown:

$$\mathrm{var}(v_j) = (\sigma_{u_j u_j})^2 I_{Q_j} - (\sigma_{u_j u_j}\rho_{\eta_j^* u_j})^2 C_j$$

where I_{Q_j} is a Q_j x Q_j identity matrix. Note also $E(\hat{\alpha} - \alpha)v_j' = \mathrm{cov}(\hat{\alpha}, v_j') = 0$ (Maddala 1983). Substituting these expressions into (A7.12) we obtain:

$$\mathrm{var}\begin{bmatrix}\hat{\beta}_j \\ \sigma_{u_j u_j}\widehat{\rho}_{\eta_j^* u_j}\end{bmatrix}$$

$$= (\sigma_{u_j u_j})^2 (G_j' G_j)^{-1} - (\sigma_{u_j u_j}\rho_{\eta_j^* u_j})^2 (G_j' G_j)^{-1}$$

$$\times\ G_j'\left[C_j - B_\alpha \mathrm{var}(\hat{\alpha}) B_\alpha'\right] G_j (G_j' G_j)^{-1} \qquad \text{(A7.13)}$$

where var $(\hat{\alpha})$ is the R x R variance/covariance matrix obtained from estimation of the multinomial logit model. The estimated variance/covariance matrix can be obtained by replacing G_j with $\hat{G}_j$, C_j with $\hat{C}_j$, B_α with $\hat{B}_\alpha$, $(\sigma_{u_j u_j} \rho_{\eta_j^* u_j})$ with $(\widehat{\sigma_{u_j u_j} \rho_{\eta_j^* u_j}})$ and $(\sigma_{u_j u_j})^2$ with $(\widehat{\sigma_{u_j u_j}})^2$ where

$$(\sigma_{u_j u_j})^2 = \frac{\sum_q (E_{jq} - X_{jq}\beta)}{Q_j}$$

It is easy to show that use of the estimated variances from OLS estimation of A7.4, ignoring the fact that α are estimated parameters, will lead to underestimation of the true variances as calculated from equation A7.13.

```
      PROGRAM HLMVCM (PARIN,OUTPUT,SHOPOUT,
     1                TAPE1=PARIN,TAPE6=OUTPUT,TAPE9=SHOPOUT)
C
C     PROGRAM:      PROGRAM TO CALCULATE CORRECTED VARIANCE/COVARIANCE
C                   MATRIX IN DISCRETE/CONTINUOUS MODEL SYSTEMS
C                   ESTIMATED BY THE TWO-STAGE HECKMAN-LEE
C                   SELECTIVITY CORRECTION METHOD WHERE THE DISCRETE
C                   MODEL IS OF THE MULTINOMIAL LOGIT FORM
C
C     DATE:         4 DECEMBER 1985
C                   (AMMENDED 24 JANUARY 1986 AND 9 MARCH 1986)
C
C     VERSION       1.02 - 1. ASSUMES Q = QJ AS FOR UNRANKED DATA
C                          2. REQUIRES USER ROUTINES
C                             VALV -  TO CALCULATE VALUES FOR THE
C                                     CONDITIONAL INDIRECT UTILITY
C                                     FUNCTIONS (CIUFS)
C                             PDERV - TO CALCULATE PARTIAL DERIVATIVES
C                                     INVOLVING THE CIUFS
C                             RECODEX - TO RECODE ANY OF THE FIRST P
C                                     COLUMN VECTORS OF THE G MATRIX
C                             VALE -  TO CALCULATE THE FUNCTION
C                                     ASSOCIATED WITH THE CONTINUOUS
C                                     CHOICE MODEL (CCM)
C
C     OPTIONS:      SEE PARIN FILE ('PROGRAM OPERATION' BLOCK)
C                   FOR MORE DETAILS
C                                                VAR     SIZE  VALUES
C                   1.  SIZE OF PRINTED MATRIX   MROW     4    2 - 20
C                       BLOCKS CAN BE ALTERED    MCOL     4    2 - 20
C                   2.  OUTPUT LEVEL CAN BE SET  VERBOSE  4    FULL,PART
C                                                              NONE
C                   3.  PROGRAM CAN TERMINATE    HOWLONG  4    VARI,DATA
C                       AFTER CALCULATING THE    OUTFORM 40    (.....)
C                       SELECTIVITY CORRECTION (SC)
C                       FACTOR AND OUTPUTING THIS
C                       WITH THE X MATRIX AND
C                       DEPENDENT VARIABLE FOR
C                       THE CCM
C
C     FORMULA:      FOR THE VARIANCE/COVARIANCE MATRIX
C
C
C
C         (-1
C         (       BETA          )     (SIGMA     ) (G G )
C         (          J          )           V V      J J
C     VAR (
C         ( SIGMA     RHO *     )                              2    T    -1
C         (       V V      N V  )        - (SIGMA     RHO *   )  (G G )
C         (        J J      J J )                V V      N V        J J
C                                                 J J      J J
C
C                                                   T            T     -1
C                                                X G    (CM) G   (G   G )
C                                                   J          J    J   J
C
C
C     WHERE
C
C     CM   =  C   - B         VAR(ALPHA) B
C              J    ALPHA                   ALPHA
C
```

```
C       REWRITE THIS, USING THE PROGRAM MATRIX VARIABLES, AS
C
C       VARB  =  CP  X  GTGINV
C
C       WHERE
C                                                      T
C       CP  =  SIGMA2  -  SIGRHO2  X  GTGINV  X  G   (CM) G
C
C       THE FOLLOWING PARAMETERS ARE SET FOR DIMENSIONS OF ARRAYS
C
C       MATRIX          ROW SIZE  COL SIZE     NAME
C
C       VAR ()            P + 1     P + 1      VARB
C       G                 Q         P + 1      G
C        J
C       CM                P + 1     P + 1      CM
C       C                 Q         Q          C  (NOTE, DIAGONAL = VECTOR)
C        J
C       B                 Q         R          B
C        ALPHA
C
C       VAR (ALPHA)       R         R          VARD
C       CP                P + 1     P + 1      CP
C         T   -1
C       (G G )            P + 1     P + 1      GTGINV
C         J J
C
C       SYMBOLIC CONSTANTS (SEE 'PARAMETER') USED IN ARRAY DIMENSIONS
C
C       Q   -   NUMBER OF OBSERVATIONS IN DCM
C       QJ  -   NUMBER OF OBSERVATIONS IN CCM FOR CHOICE J
C       P   -   NUMBER OF PARAMETERS FOR CCM
C       R   -   NUMBER OF PARAMETERS FOR DCM
C       NA  -   NUMBER OF ALTERNATIVES IN DCM FOR EACH INDIVIDUAL
C       NP  -   NUMBER OF VARIABLES IN DCM
C       (T  -   USED AS A SUPERSCRIPT REFERS TO MATRIX TRANSPOSE)
C
C       NOTE. IN MOST MATRICES, AN ACTUAL NUMBER OF ELEMENTS DEFINED FOR
C             A PARTICULAR MODEL, IS USED IN THE CALCULATION (SEE TC,
C             NAC, ETC). HOWEVER, THE MAX. DIMENSION MUST STILL BE PASSED
C             AS A PARAMETER TO DEFINE THE ADJUSTABLE SIZED MATRICES.
C       NOTE. AS MENTIONED ABOVE, QJ, THE ACTUAL NUMBER OF PERSONS
C             CHOOSING A GIVEN ALTERNATIVE J (WITH QJ .LE. Q), SHOULD
C             BE USED IN CCM CALCULATIONS. AT PRESENT IT IS
C             IMPLICITELY EQUAL TO Q.
C
C       THE IMSL LIBRARY ROUTINES ARE USED TO CALCULATE MATRIX INVERSE
C                     T   -1
C       GTGINV  =  (G G )   OF MATRIX G
C                   J J               J
C
C       AND THE CUMULATIVE NORMAL DISTRIBUTION FUNCTION INVERSE
C
C                      -1
C       J (D)  =  PHI    (D)
C
C       FILES:         PARIN   - INPUT DATA; MATRICES, VECTORS, ETC
C                      OUTPUT  - INTERACTIVE OUTPUT
C                      CCMOUT  - OUTPUT DATA; G MATRIX
C
C       PARIN FILE - INPUT PARAMETERS
C
C       NOTE. THE NUMBERED CODE SYSTEM USED IN THE PARIN FILE SERVES
```

```
C          TWO PURPOSES. FIRSTLY, IT ENABLES DATA IN THE PARIN FILE
C          TO BE EASILY PINPOINTED. SECONDLY, THESE CODES ARE
C          USED AS COMMENTS IN THE PROGRAM TO ALLOW EASY IDENTIFICATION
C          OF THE SECTIONS OF DATA BEING MANIPULATED. THE FORMAT
C          INFORMATION BELOW SPCIFIES HOW THE PROGRAM READS IN THE
C          DATA FROM THE PARIN FILE. '+ COMMENT' SIGNIFIES THAT
C          THE USER CAN PROVIDE ANY COMMENT. THESE COMMENTS ARE
C          NOT USED BY THE PROGRAM.
C
C      LINE         FORMAT      CODE + ITEMS
C
C       1           I2          '00' + 'PROGRAM OPERATION DATA'
C       2           2X,2I4,            MROW - MAX ROW PRINT BLOCK
C                                             (A GOOD DEFAULT VALUE IS 12)
C                   3(A,2X)            MCOL - MAX COL PRINT BLOCK
C                                             (A GOOD DEFAULT VALUE IS 8)
C                                      VERBOSE - OUTPUT PRINT LEVEL
C                                      HOWLONG - TERMINATE AFTER CALCULATION
C                                                OF SC TERM
C                                      OUTFORM - FORMAT TO WRITE TAPE 9
C
C       3           I2          '10' + 'DISCRETE CHOICE MODEL DATA'
C       4           I2          '11' + COMMENT
C       5           2X,I4,A            R - NUMBER OF PARAMETERS IN DCM
C                                      FORM - DATA FORMAT OF NEXT LINE
C       6           FORM               ALPHA(R) VALUES
C                                      IE. R COLUMNS TO BE READ IN
C       7           I2          '12' + COMMENT
C       8           2X,I4,A            R - NUMBER OF PARAMETERS IN DCM
C                                      FORM - DATA FORMAT OF NEXT R LINES
C       9(+R)       FORM               VAR(ALPHA)(R,R) VALUES
C                                      (VARIANCE/COVARIANCE MATRIX OF DCM)
C                                      IE. R ROWS OF R COLUMNS TO BE READ IN
C       9+R         I2          '13' + COMMENT
C       10+R        2X,3I4,A           Q  - NUMBER OF PERSONS IN DCM
C                                      NA - NUMBER OF ALTERNATIVES IN
C                                           THE CHOICE SET OF EACH PERSON
C                                           (ASSUMED CONSTANT)
C                                      NP - NUMBER OF VARIABLES IN DCM
C                                      FORM - DATA FORMAT OF NEXT Q LINES
C       11+R(+Q)    FORM               CHOSALT(Q), Z(Q,NP) VALUES
C                                      IE. Q ROWS OF 1 + NP X NA COLUMNS TO
C                                      BE READ IN INCLUDING A CHOSEN ALTERNATIVE
C                                      SELECTOR
C
C       11+R+Q      I2          '20' + 'CONTINUOUS CHOICE MODEL DATA'
'21' + COMMENT
C      Q2 - NUMBER OF PERSONS IN CCM ASSOCIATED
C                                           WITH CHOICE J
C                                      P - NUMBER OF VARIABLES IN CCM
C                                      FORM - DATA FORMAT OF NEXT QJ LINES
C       3+Q)      FORM               CONTAINING THE X(QJ,P) AND E(QJ) VALUES
C                                      (NOTE THAT G = (X:Y) WHERE
C                                      Y(Q) = VECTOR OF TAYLOR EXPANSION TERMS)
C                                      IE. QJ ROWS OF P+1 COLUMNS TO BE READ IN
C       3Q        I2          '22' + COMMENT
C       4Q        2X,A               FORM - DATA FORMAT OF NEXT LINE
C       5Q        FORM               SIGRHO2 - THE VARIANCE * COVARIANCE TERM
C                                      (I.E. THE PARAMETER ESTIMATE ASSOCIATED
C                                      WITH THE SC FACTOR)
C       6Q        I2          '23' + COMMENT
C       7Q        2X,I4,A            P - NUMBER OF PARAMETERS IN CCM
C                                      FORM - DATA FORMAT OF NEXT LINE
```

```
C       8Q        FORM                 BETA(P) VALUES
C                                         IE. P COLUMNS TO BE READ IN
C       9Q        I2           '30' + COMMENT (END OF DATA)
C
C       20+R+2Q     - - - - -  TOTAL LINES IN FILE
C
      CHARACTER    VERBOSE*4, HOWLONG*4, TASK*30, MODEL*20, FORM*40,
     1             OUTFORM*40
      INTEGER      Q, P, R, NA, NP, QC, PC, RC, QI, PI, RI, MROW, MCOL
      REAL         J
C
      PARAMETER    (Q = 300, P = 8, R = 9, NA = 10, NP = 9)
C
      DIMENSION    G(Q,P+1), GW(P+1,P+1), VW(P+1), GTGINV(P+1,P+1),
     1             C(Q), B(Q,R), VARD(R,R), T1(P+1,R), T2(R,P+1),
     1             TEM(R), CP(P+1,P+1), VARB(P+1,P+1), BETA(P), E(Q),
     1             Z(NA,NP), ALPHA(R), D(Q), J(Q), PHI(Q), DA(NA),
     1             CSEB(P+1),CTB(P+1),SUMS(P+2)
C
      COMMON  /REPC/  VERBOSE
      COMMON  /REPD/  MROW, MCOL
C
      REWIND (1,ERR = 920)
      LNO = -1
C
C     READ IN PROGRAM OPERATION DATA
C
C     '00'
C
      READ (1,1000,END=931,ERR=931) LNO,MROW,MCOL,VERBOSE,HOWLONG,
     1                              OUTFORM
 1000 FORMAT (I2/2X,2I4,3(A,2X))
      IF (LNO.NE.00) GOTO 931
      IF (MROW.GT.20 .OR. MROW.LT.2) MROW = 12
      IF (MCOL.GT.20 .OR. MCOL.LT.2) MCOL = 8
      IF (VERBOSE.NE.'FULL' .AND. VERBOSE.NE.'PART' .AND.
     1    VERBOSE.NE.'NONE') VERBOSE = 'FULL'
      IF (HOWLONG.NE.'VARI' .AND. HOWLONG.NE.'DATA') HOWLONG = 'VARI'
C
C     OUTPUT A HEADER
C
      MODEL = 'MAXIMUM EXPENDITURE'
      TASK = 'CALCULATE VARIANCE MATRIX'
      IF (HOWLONG.EQ.'DATA') TASK = 'OUTPUT DATA FOR CCM INC. SC TERM'
      WRITE (6,6000) MODEL,MODEL,VERBOSE,TASK
 6000 FORMAT (//2X,'PROGRAM HLMVCM  -  CALCULATION OF CORRECTED',
     1        ' VARIANCE/COVARIANCE MATRIX WHEN USING THE',
     1        ' HECKMAN/LEE SELECTIVITY CORRECTION METHOD'/
     1        2X,81('=')//21X,'USES    -  DISCRETE CHOICE MODEL - ',A/
     1        28X,'-  CONTINUOUS CHOICE MODEL - ',A//21X,'OUTPUT -  ',A/
     1        21X,'TASK    -  ',A//21X,41('=')////)
C
C     READ IN DISCRETE CHOICE MODEL DATA - ALPHA, VARD, Z
C     AND CALCULATE THE DATA DEPENDENT MATRICES - D, J, PHI, B, C
C
C     '10'
C
      READ (1,1010,END=931,ERR=931) LNO
 1010 FORMAT (I2)
      IF (LNO.NE.10) GOTO 931
C
C     '11'
      READ (1,1011,END=931,ERR=931) LNO,RC,FORM
```

```
 1011 FORMAT (I2/2X,I4,A)
      IF (LNO.NE.11) GOTO 931
      IF (RC.GT.R) GOTO 932
      READ (1,FORM,END=931,ERR=911) (ALPHA(L),L=1,RC)
      CALL PRINTM (ALPHA,1,R,1,RC,'DCM CHOICE VECTOR  ALPHA',2)
C
C     '12'
      READ (1,1012,END=931,ERR=931) LNO,RI,FORM
 1012 FORMAT (I2/2X,I4,A)
      IF (LNO.NE.12) GOTO 931
      IF (RI.NE.RC) GOTO 932
      DO 20 K = 1, RC
          READ (1,FORM,END=931,ERR=912) (VARD(K,L),L=1,RC)
 20   CONTINUE
      CALL PRINTM (VARD,R,R,RC,RC,'EST''D VAR/COV MATRIX  VARD',2)
C
C     '13'
      CALL LSCM (Q,R,NA,NP,QC,RC,ALPHA,Z,B,D,J,PHI,DA,IERR)
      IF (IERR.EQ.1) GOTO 931
      LNO = 13
      IF (IERR.EQ.2) GOTO 932
      IF (IERR.EQ.3) GOTO 913
      IF (IERR.EQ.4) GOTO 914
C
C     CALCULATE THE C DATA VECTOR
C
      CALL DEFINEC (J,PHI,Q,QC,C)
C
C     READ IN CONTINUOUS CHOICE MODEL DATA - X/G, E, SIGRHO2, BETA
C     AND CALCULATE THE DATA DEPENDENT MATRICES - G, SIGMA2
C
C     '20'
C
      READ (1,1020,END=931,ERR=931) LNO
 1020 FORMAT (I2)
      IF (LNO.NE.20) GOTO 931
C
C     '21'
      READ (1,1021,END=931,ERR=931) LNO,QI,PC,FORM
 1021 FORMAT (I2/2X,2I4,A)
      IF (LNO.NE.21) GOTO 931
      IF (QI.NE.QC) GOTO 932
      IF (PC.GT.P) GOTO 932
      DO 40 K = 1, QC
          READ (1,FORM,END=931,ERR=921) (G(K,L),L=1,PC),E(K)
 40   CONTINUE
      CALL PRINTM (E,1,Q,1,QC,'CONT. CHOICE MODEL VECTOR  E',3)
C
C     CALCULATE THE REST OF THE G DATA MATRIX
C
      CALL RESTOFG (PHI,Q,P,QC,PC,G)
C
C     WRITE CHOICE DATA AND TERMINATE PROGRAM IF REQUESTED
C
      IF (HOWLONG.EQ.'DATA') THEN
          REWIND 9
          DO 60 K = 1, QC
              WRITE (9,OUTFORM,ERR=940) (G(K,L),L=1,PC),E(K),G(K,PC+1)
 60       CONTINUE
          READ (1,1030,END=931,ERR=931) LNO
          IF (LNO.NE.30) GOTO 931
          GOTO 990
      ENDIF
```

```
C
C     '22'
      READ (1,1022,END=931,ERR=931) LNO,FORM
 1022 FORMAT (I2/2X,A)
      IF (LNO.NE.22) GOTO 931
      READ (1,FORM,END=931,ERR=922) SIGRHO2
      IF (VERBOSE.EQ.'FULL') WRITE (6,6022) SIGRHO2
 6022 FORMAT (//'  SIGRHO2  =  ',F12.5/'  ======='//)
C
C     '23'
      READ (1,1023,END=931,ERR=931) LNO,PI,FORM
 1023 FORMAT (I2/2X,I4,A)
      IF (LNO.NE.23) GOTO 931
      IF (PI.NE.PC) GOTO 932
      READ (1,FORM,END=931,ERR=923) (BETA(L),L=1,PC)
      CALL PRINTM (BETA,1,P,1,PC,'CONT. CHOICE MODEL VECTOR  BETA',1)
C
C     '30'
C
      READ (1,1030,END=931,ERR=931) LNO
 1030 FORMAT (I2)
      IF (LNO.NE.30) GOTO 931
C
C     PRODUCE/PRINT MEANS FOR OLS REGRESSION COMPARISONS
C
      IF (VERBOSE.NE.'NONE') THEN
          DO 520 K = 1, PC+1
              SUMM = 0.0
              DO 500 I = 1, QC
                  SUMM = SUMM + G(I,K)
 500          CONTINUE
              SUMS(K) = SUMM / QC
 520      CONTINUE
          SUMM = 0.0
          DO 540 I = 1, QC
              SUMM = SUMM + E(I)
 540      CONTINUE
          SUMS(PC+2) = SUMM / QC
          WRITE (6,6560) (SUMS(I),I=1,PC+2)
 6560     FORMAT (' OUTPUT X VECTOR MEANS FOR OLS REGRESSION',
     1               ' COMPARISON'//5X,8F9.4/)
      ENDIF
C
C     PERFORM THE VARIOUS MATRIX OPERATIONS ASSOCIATED WITH
C     THE CALCULATION OF THE VARIANCE/COVARIANCE MATRIX - VARB -
C
C     CALCULATE THE ESTIMATED SIGMA SQUARE VALUE -  SIGMA2
C
      CALL SIGMA (G,E,BETA,Q,P,QC,PC,SIGMA2)
C
C     CALCULATE THE GTGINV MATRIX
C
      CALL GETINV (G,Q,P+1,QC,PC+1,GW,GTGINV,VW)
C
C
C     CALCULATE THE (TEMPORARY) MATRIX  -  CP  =  T1  VAR(ALPHA)  T2
C
C                       T
C     WHERE  T1   =  G    B
C
C                       T
C            T2   =  B    G
C
```

```
      CALL MATMUL (G,B,Q,P+1,R,QC,PC+1,RC,T1)
      CALL MATMUL (B,G,Q,R,P+1,QC,RC,PC+1,T2)
      DO 680 K = 1, PC+1
          DO 620 L = 1, RC
              TEMP = 0.0
              DO 600 I = 1, RC
                  TEMP = TEMP + T1(K,I) * VARD(I,L)
 600          CONTINUE
              TEM(L) = TEMP
 620      CONTINUE
          DO 660 L = 1, PC+1
              TEMP = 0.0
              DO 640 I = 1, RC
                  TEMP = TEMP + TEM(I) * T2(I,L)
 640          CONTINUE
              CP(K,L) = TEMP
 660      CONTINUE
 680  CONTINUE
C                                                 T
C     CALCULATE THE (TEMPORARY) MATRIX - VARB = G  (CM) G
C              T                                  J        J
C     VARB = G  C  G  -  CP
C
      DO 720 K = 1, PC+1
          DO 720 L = 1, PC+1
              TEMP = 0.0
              DO 700 I = 1, QC
                  TEMP = TEMP + G(I,K) * C(I) * G(I,L)
 700          CONTINUE
              VARB(K,L) = TEMP - CP(K,L)
 720  CONTINUE
C
C     CALCULATE THE MATRIX -  CP
C     CP  =  SIGMA2  -  SIGRHO2  X  GTGINV  X  VARB
C
      DO 760 K = 1, PC+1
          DO 760 L = 1, PC+1
              TEMP = 0.0
              DO 740 I = 1, PC+1
                  TEMP = TEMP + GTGINV(K,I) * VARB(I,L)
 740          CONTINUE
              IF (K.EQ.L) THEN
                  CP(K,L) = SIGMA2 - SIGRHO2 * TEMP
              ELSE
                  CP(K,L) =        - SIGRHO2 * TEMP
              ENDIF
 760  CONTINUE
C
C     CALCULATE THE MATRIX -  VARB
C     VARB  =  CP  X  GTGINV
C
      DO 800 K = 1, PC+1
          DO 800 L = 1, PC+1
              TEMP = 0.0
              DO 780 I = 1, PC+1
                  TEMP = TEMP + CP(K,I) * GTGINV(I,L)
 780          CONTINUE
              VARB(K,L) = TEMP
 800  CONTINUE
C
C     CALCULATE THE MATRIX - CSEB - CORRECTED STANDARD ERROR
C     CSEB  =  SQRT ( DIAGONAL ELEMENTS OF VARB )
C     AND - CTB - CORRECTED T STATISTIC ON BETA AND SIGRHO2
```

```
C     CTB  =  BETA / CSEB      FOR 1 .. PC
C     CTB  =  SIGRHO2 / CSEB   FOR PC+1
C
      DO 820 K = 1, PC+1
          CSEB(K) = SQRT (VARB(K,K))
 820  CONTINUE
      DO 840 K = 1, PC
          CTB(K) = BETA(K) / CSEB(K)
 840  CONTINUE
      CTB(PC+1)  = SIGRHO2 / CSEB(PC+1)
C
C     OUTPUT - VARB, CSEB, CTB - AND FINISH
C
      CALL PRINTM (VARB,P+1,P+1,PC+1,PC+1,
     1             'VARIANCE / COVARIANCE MATRIX',1)
      CALL PRINTM (CSEB,1,P+1,1,PC+1,
     1             'CORRECTED STANDARD ERROR VECTOR',1)
      CALL PRINTM (CTB,1,P+1,1,PC+1,
     1             'CORRECTED ''T'' STATISTIC VECTOR',1)
      GOTO 990
C
C     FINISH/ERROR MESSAGES
C
 911  WRITE (6,6911)
 6911 FORMAT (' ERROR READING ALPHA VECTOR DATA')
      GOTO 990
 912  WRITE (6,6912)
 6912 FORMAT (' ERROR READING VAR(ALPHA) MATRIX DATA')
      GOTO 990
 913  WRITE (6,6913)
 6913 FORMAT (' ERROR READING Z MATRIX DATA')
      GOTO 990
 914  WRITE (6,6914)
 6914 FORMAT (' ERROR (LSCM)  ..  COMPUTING J VECTOR')
      GOTO 990
 921  WRITE (6,6921)
 6921 FORMAT (' ERROR READING G/E MATRIX DATA')
      GOTO 990
 922  WRITE (6,6922)
 6922 FORMAT (' ERROR READING SIGRHO2 DATA')
      GOTO 990
 923  WRITE (6,6923)
 6923 FORMAT (' ERROR READING BETA VECTOR DATA')
      GOTO 990
 920  WRITE (6,6920)
 6920 FORMAT (' NO DATA FILE')
      GOTO 990
 931  WRITE (6,6931) LNO
 6931 FORMAT (' ERROR READING DATA FILE AFTER LAST CODE = ',I2)
      GOTO 990
 932  WRITE (6,6932) LNO
 6932 FORMAT (' MISMATCHED PARAMETERS AFTER CODE = ',I2)
      GOTO 990
 940  WRITE (6,6940)
 6940 FORMAT (' ERROR WRITING G/E MATRIX DATA')
C
 990  WRITE (6,6990)
 6990 FORMAT (//' END HLMVCM'/)
C
      END
```

```
C
      SUBROUTINE LSCM (Q,R,NA,NP,QC,RC,ALPHA,Z,B,D,J,PHI,DA,IERR)
C
C     ROUTINE TO CALCULATE THE CERTAIN VALUES FROM THE LOGIT DCM (LSCM)
C
C     1.  READ IN THE MATRIX - Z(Q,NP)
C     2.  CALCULATE THE VECTORS - D, J(D) AND PHI(J)
C
C     D(ALPHA,Z  )  =  EXP  ( V(ALPHA,Z   ) )  /  SUM  ( EXP  ( V(ALPHA,Z   ) ) )
C              JQ           (            JQ  )        I   (      (            IQ  ) )
C
C     WHERE EXP   =  EXPONENTIAL FUNCTION
C           V()   =  USER SPECIFIED UTILITY FUNCTION, SEE - VALV -
C           SUM   =  SUMMATION FUNCTION
C
C                            -1  (                )
C     J(ALPHA,Z  )  =  CND       ( D(ALPHA,Z  )  )
C              JQ                (          JQ   )
C
C              -1
C     WHERE CND    =  INVERSE OF CUMULATIVE NORMAL DISTRIBUTION
C
C     PHI  ( J(ALPHA,Z   ) )  =  DND  ( J(ALPHA,Z   ) )  /  D(ALPHA,Z  )
C          (           JQ  )          (           JQ  )              JQ
C
C     WHERE DND   =  DENSITY OF CUMULATIVE NORMAL DISTRIBUTION
C           CND   =  STANDARD NORMAL DISTRIBUTION
C
C     3.  CALCULATE THE MATRIX B, USING TERMS FROM FUNCTION - PDERV -
C
C                        ( DND(J(ALPHA,Z    )) )
C                        (                JQ   )
C                        (                  J  )
C     B(Q,R)    =   DEL  ( ------------------  )  /  DEL  (ALPHA )
C                        (   D(ALPHA,Z    )    )                 R
C                        (              JQ     )
C                        (                J    )
C
C     DA   =  A SCRATCH VECTOR USED TO STORE VALUES OF D FOR ALL ALT'S
C
C     THE DIMENSIONS ARE -
C
C     MATRIX       ROW   COL
C
C     Z            NA    NP   (NOTE. SPECIAL USE OF Z)
C     ALPHA        -     R
C     D            -     Q
C     J            -     Q
C     PHI          -     Q
C     B            Q     R
C     DA           -     NP
C
C     PARIN FILE - INPUT PARAMETERS (SEE RELEVANT SECTION IN MAINLINE)
C
C     CODE - '13'
C     ITEM - CHOSALT(Q), Z(Q,NP) VALUES
C            Q ROWS OF 1 + R X N COLUMNS TO BE READ IN
C            INCLUDING A CHOSEN ALTERNATIVE SELECTOR
C
C     NOTE - HOWEVER THAT THE Z MATRIX READ IN IS NOT Z(Q,NP) AS
C            ORIGINALLY DEFINED. INSTEAD WE ARE READING IN THE DATA
```

```
C           ONE LINE AT A TIME AND CALCULATING RESULTS PROGRESSIVELY.
C           THE Z(NA,NP) IN FACT STORES NA ALTERNATIVE SETS, FROM THE
C           ONE INPUT LINE, WHICH WILL BE DEPENDANT ON THE DATA FILE.
C           THE MAIN REASON FOR INCLUDING Z (AS A WORK ARRAY) IN THE
C           SUBROUTINE PARAMETER LIST IS SO THAT IT CAN HAVE
C           ADJUSTABLE DIMENSIONS TO MATCH ALPHA.
C
C     IMSL ROUTINE - MDNRIS - IS USED TO CALCULATE THE INVERSE OF
C                             THE CUMULATIVE NORMAL DISTRIBUTION
C     USER ROUTINE - VALV -   IS USED TO PROVIDE THE EXACT FORM OF
C                             THE CIUFS
C     USER ROUTINE - PDERV -  IS USED TO PROVIDE THE PARTIAL
C                             DERIVATIVES OF THE CIUFS
C
C     ERROR RETURNS, VIA - IERR -, ARE -
C
C     IERR = 0 - ALL OK
C            1 - ERROR READING DATA FILE (C/F EXIT 931 IN MAINLINE)
C            2 - MISMATCHED PARAMETERS   ( "    "   932 "      "    )
C            3 - ERROR READING Z MATRIX  ( "    "   913 "      "    )
C            4 - IMSL ERROR (MDNRIS)     ( "    "   914 "      "    )
C
      CHARACTER   FORM*40
      INTEGER     Q, R, NA, NP, QC, RC, NAC, NPC, IERR, CHOSALT
      REAL        Z, ALPHA, D, J, PHI, B, DA, VALV, PDERV
C
      DIMENSION   Z(NA,NP), ALPHA(R), D(Q), J(Q), PHI(Q), B(Q,R), DA(NA)
C
      PI = 3.14159265
      CON = 1.0 / SQRT (2.0 * PI)
C
C     READ IN Z MATRIX, A PERSON AT A TIME AND CALCULATE THE
C     D, J, PHI AND B MATRICES, USING IMSL ROUTINE - MDNRIS
C
      READ (1,1013,END=910,ERR=910) LNO,QC,NAC,NPC,FORM
 1013 FORMAT (I2/2X,3I4,A)
      IF (LNO.NE.13) GOTO 910
      IF (QC.GT.Q) GOTO 920
      IF (NAC.GT.NA) GOTO 920
      IF (NPC.GT.NP) GOTO 920
      IMEM = 0
      DO 100 K = 1, QC
          READ (1,FORM,END=910,ERR=930) CHOSALT,
     1                                  ((Z(I,L),L=1,NPC),I=1,NAC)
          CALL PRINTM (Z,NA,NP,NAC,NPC,'DISC. CHOICE MOD. MATRIX  Z',3)
          PSUM = 0.0
          DO 20 I = 1, NAC
              DA(I) = EXP(VALV(I,K,ALPHA,Z,R,NA,NP,RC,NAC,NPC))
              PSUM = PSUM + DA(I)
 20       CONTINUE
          CPROB = 0.0
          DO 40 I = 1, NAC
              DA(I) = DA(I) / PSUM
              IF (I.EQ.CHOSALT) CPROB = DA(I)
 40       CONTINUE
          D(K) = CPROB
          CALL MDNRIS (CPROB,VALJ,IERR)
          IF (IERR.NE.0) THEN
              IMEM = 1
              J(K) = 0.0
              PHI(K) = 0.0
              WRITE (6,6040) IERR
 6040         FORMAT (' IMSL ERROR (MDNRIS)  ..  IER = ',I3,
```

```
     1                   ' ROUTINE WILL CONTINUE')
          ELSE
              J(K) = VALJ
              PHI(K) = CON * EXP(-0.5 * VALJ**2) / CPROB
          ENDIF
          VALJ = -1.0 * (J(K) + PHI(K))
          DO 100 L = 1, RC
              T1 = 0.0
              T2 = 0.0
              DO 60 I = 1, NAC
                  T3 = PDERV (I,K,L,ALPHA,Z,R,NA,NP,RC,NAC,NPC)
                  IF (I.EQ.CHOSALT) T2 = T3
                  T1 = T1 + DA(I) * T3
 60           CONTINUE
              B(K,L) = VALJ * (T2 - T1)
 100  CONTINUE
      IF (IMEM.NE.0) GOTO 940
      CALL PRINTM (D,1,Q,1,QC,'PROBABILITY VECTOR  D',2)
      CALL PRINTM (J,1,Q,1,QC,'J  =  INVERSE NORMAL OF D',3)
      CALL PRINTM (PHI,1,Q,1,QC,'DENSITY FUNCTION VECTOR  PHI',2)
      CALL PRINTM (B,Q,R,QC,RC,'TAYLOR COEFFICIENT MATRIX  B',3)
C
 900  IERR = 0
      RETURN
 910  IERR = 1
      RETURN
 920  IERR = 2
      RETURN
 930  IERR = 3
      RETURN
 940  IERR = 4
      RETURN
C
      END
```

```
C
      SUBROUTINE DEFINEC (J,PHI,Q,QC,C)
C
C     ROUTINE TO CALCULATE THE DIAGONAL MATRIX C, GIVEN BY TERMS
C     CALCULATED IN SUBROUTINE - LSCM -
C
C                 DND(J(ALPHA,Z    ))  ( DND(J(ALPHA,Z    ))                  )
C                              JQ      (              JQ                      )
C                                J     (                J      J(ALPHA,Z    ) )
C     C(Q,Q)  =  -------------------   ( ------------------ -            JQ   )
C                   D(ALPHA,Z    )     (    D(ALPHA,Z    )                  J )
C                             JQ       (              JQ                      )
C                               J      (                J                     )
C
C     WHERE DND  =  DENSITY OF CUMULATIVE NORMAL DISTRIBUTION
C           CND  =  STANDARD NORMAL DISTRIBUTION
C
C     THE DIMENSIONS ARE -
C
C     MATRIX  ROW     COL
C
C     J        -       Q
C     PHI      -       Q
C     C        Q       Q          - ALTHOUGH WE ONLY WANT THE DIAGONAL
C                                     AND ACTUALLY DEFINE A VECTOR C(Q)
C
      INTEGER Q, QC
      REAL    J, PHI, C
C
      DIMENSION   J(Q), PHI(Q), C(Q)
C
      DO 100 I = 1, QC
          C(I) = PHI(I) * (PHI(I) - J(I))
  100 CONTINUE
      CALL PRINTM (C,1,Q,1,QC,'DIAGONAL MATRIX  C',3)
C
      RETURN
C
      END
```

```
C
      SUBROUTINE RESTOFG (PHI,Q,P,QC,PC,G)
C
C     ROUTINE TO CALCULATE THE LAST COLUMN OF G, GIVEN BY PHI(Q)
C     CALCULATED IN SUBROUTINE - LSCM -
C
C                     -DND(J(ALPHA,Z    ))
C                                   JQ
C                                     J
C     G(Q,P+1)   =   -------------------   =  SCLEE FACTOR
C                        D(ALPHA,Z    )
C                                 JQ
C                                   J
C
C     WHERE DND   =  DENSITY OF CUMULATIVE NORMAL DISTRIBUTION
C           CND   =  STANDARD NORMAL DISTRIBUTION
C
C     THE DIMENSIONS OF G ARE -
C
C     MATRIX  ROW     COL
C
C     G       Q       P+1
C
      INTEGER Q, P, QC, PC
      REAL    PHI, G
C
      DIMENSION   PHI(Q), G(Q,P+1)
C
      CALL RECODEX (G,Q,P,QC,PC)
      DO 20 I = 1, QC
          G(I,PC+1) = -1.0 * PHI(I)
 20   CONTINUE
      CALL PRINTM (G,Q,P+1,QC,PC+1,'G MATRIX - I.E. CCM DATA MATRIX WITH
     1 SC TERM',3)
C
      RETURN
C
      END
```

```
C
      SUBROUTINE SIGMA (G,E,BETA,Q,P,QC,PC,SIGMA2)
C
C     ROUTINE TO CALCULATE THE VARIANCE FACTOR - SIGMA2
C
C                  2                                       2
C     ( SIGMA      )               (
C     (        V V )  =  SUM       ( E    - BETA  X    )  /  Q
C     (         J J)       Q       (  JQ             JQ )       J
C
C     WHERE SUM   =  SUMMATION FUNCTION
C
C     THE DIMENSIONS ARE -
C
C     MATRIX   ROW      COL
C
C     G        Q        P+1
C     E        -        Q
C     BETA     -        P
C
C     USER ROUTINE - VALE - IS USED TO PROVIDE THE EXACT FORM OF
C                           THE EXPENDITURE FUNCTION
C
      CHARACTER VERBOSE*4
      INTEGER   Q, P, QC, PC
      REAL      G, E, BETA, SIGMA2, VALE
C
      DIMENSION G(Q,P+1), E(Q), BETA(P)
C
      COMMON  /REPC/  VERBOSE
C
      SIGMA2 = 0.0
      DO 100 K = 1, QC
          SIGMA2 = SIGMA2 + (E(K) - VALE(K,BETA,G,Q,P,QC,PC))**2
 100  CONTINUE
      SIGMA2 = SIGMA2 / QC
      IF (VERBOSE.NE.'NONE') WRITE (6,6100) SIGMA2
 6100 FORMAT (//'  SIGMA2   =   ',F12.5/'  ======'//)
C
      RETURN
C
      END
```

```
C
      SUBROUTINE MATMUL (M1,M2,RR,C1,C2,RRC,C1C,C2C,M3)
C
C     ROUTINE TO MULTIPLY TWO REAL MATRICES - M1, M2 - TO GIVE M3
C
C                   T
C         M3   =  M1   X   M2
C
C     THE DIMENSIONS ARE -
C
C     MATRIX   ROW      COL
C
C     M1       RR       C1
C     M2       RR       C2
C     M3       C1       C2
C
      INTEGER RR, C1, C2, RRC, C1C, C2C
      REAL    M1, M2, M3
C
      DIMENSION   M1(RR,C1), M2(RR,C2), M3(C1,C2)
C
      DO 200 I = 1, C1C
          DO 200 J = 1, C2C
              TEMP = 0.0
              DO 100 K = 1, RRC
                  TEMP = TEMP + M1(K,I) * M2(K,J)
 100          CONTINUE
              M3(I,J) = TEMP
 200  CONTINUE
C
      RETURN
C
      END
```

```
C
      SUBROUTINE GETINV (G,M,C,MC,CC,GW,GI,VW)
C
C     ROUTINE TO CALCULATE THE INVERSE MATRIX - GI FROM G, WHERE
C
C                     T   -1
C        GI   =   ( G G )
C
C     NOTE THE USE OF WORKING ARRAYS - GW, VW (REQUIRED BY IMSL)
C
C     IMSL ROUTINE - LINV1F - IS USED TO CALCULATE THE INVERSE OF
C                             THE MATRIX PRODUCT GW
C                             HENCE IT IS NECESSARY TO USE THE
C                             CORRECT ROW DIMENSIONS
C
C     THE DIMENSIONS ARE -
C
C     MATRIX  ROW     COL
C
C     G       M       C
C     GW      C       C
C     VW      -       C
C     GI      C       C
C
      INTEGER M, C, MC, CC
      REAL    G, GW, VW, GI
C
      DIMENSION   G(M,C), GW(C,C), VW(C), GI(C,C)
C
      CALL MATMUL (G,G,M,C,C,MC,CC,CC,GW)
      CALL PRINTM (GW,C,C,CC,CC,'G (TRANSPOSE)  X  G  MATRIX',3)
C
      IDGT = 3
      CALL LINV1F (GW,CC,C,GI,IDGT,VW,IERR)
      IF (IERR.NE.0) WRITE (6,6000) IERR
 6000 FORMAT (' IMSL ERROR (LINV1F)  ..  IER = ',I3,
     1        ' ROUTINE WILL CONTINUE')
      CALL PRINTM (GI,C,C,CC,CC,'IMSL INVERSE  G (T) G  MATRIX',2)
C
      RETURN
C
      END
```

```
C
      SUBROUTINE PRINTM (M,R,C,RC,CC,HEAD,LEVEL)
C
C     ROUTINE TO PRINT ANY MATRIX, BY SECTIONS, WITH A HEADER
C     FOR A GIVEN PRIORITY LEVEL
C
C     THE DIMENSIONS OF M ARE -
C
C     MATRIX  ROW     COL
C
C     M       R       C
C
C     THE HEADER IS CONTAINED IN 'HEAD'
C
C     THE PRIORITY LEVELS ARE -
C
C     1  -  ALWAYS PRINTED
C     2  -  PRINTED IF 'VERBOSE' IS PART/FULL
C     3  -     "     "       "      "  FULL
C
      CHARACTER   HEAD*(*), VERBOSE*4, FORM*160
      INTEGER     R, C, RC, CC, MROW, MCOL
      REAL        M
C
      DIMENSION   M(R,C)
C
      COMMON  /REPC/  VERBOSE
      COMMON  /REPD/  MROW, MCOL
C
      IF (VERBOSE.EQ.'PART' .AND. LEVEL.EQ.3 .OR.
     1    VERBOSE.EQ.'NONE' .AND. LEVEL.NE.1) GOTO 900
      DO 20 I = LEN(HEAD), 1, -1
          IF (HEAD(I:I).NE.' ') THEN
              LENGTH = I
              GOTO 100
          ENDIF
 20   CONTINUE
      LENGTH = 0
 100  IF (LENGTH.GT.40) LENGTH = 40
      WRITE (FORM,6100) LENGTH,LENGTH
 6100 FORMAT ('(//2X,A',I2,',''  ('',I2,'' - '',I2,'') X ('',I2',
     1        ','' - '',I2,'')''/2X,',I2,'(''='')//8X,10I10)')
C
      DO 300 I = 1, RC, MROW
          IIM = I + MROW - 1
          IF (IIM.GT.RC) IIM = RC
          DO 300 J = 1, CC, MCOL
              JJM = J + MCOL - 1
              IF (JJM.GT.CC) JJM = CC
              WRITE (6,FORM) HEAD,I,IIM,J,JJM,(JJ,JJ=J,JJM)
              DO 200 II = I, IIM
                  WRITE (6,6101) II,(M(II,JJ),JJ=J,JJM)
 6101             FORMAT (/I10,3X,10(F10.4))
 200          CONTINUE
 300  CONTINUE
      WRITE (6,'(/)')
C
 900  RETURN
C
      END
```

```
C
      REAL FUNCTION VALV (I,K,ALPHA,Z,R,NA,NP,RC,NAC,NPC)
C
C     ROUTINE TO CALCULATE THE COND. INDIRECT UTILITY FUNCTION TERMS -
C
C     VALV(I,K)   =   V(ALPHA,Z   )
C                              IK
C
C     I   =   1 - N, REFERS TO AN ALTERNATIVE (CHOSEN OR NOT)
C     K   =   1 - Q, REFERS TO AN INDIVIDUAL FROM THE SET OF INDIVIDUALS
C
C     WHERE
C
C     V( )   =  IS A COND. INDIRECT UTILITY FUNCTION COMPRISING THE
C               PARAMETER VECTOR - ALPHA - AND VARIABLE VECTOR - Z   -
C                                                                 IK
C     ALPHA   =   (ALPHA , .... , ALPHA )
C                       1              R
C
C     Z     =   (Z     , .... , Z     ), IS AN 'NP' VECTOR
C      IK         IK              IK
C                   1               NP
C
C     THIS ROUTINE MUST BE SUPPLIED BY THE USER FOR A
C     SPECIFIC FORM OF CIUF - V() -
C     IT IS SUPPLIED WITH THE VECTORS - ALPHA - AND - Z -
C     NOTE. AT PRESENT, IT IS ONLY DEPENDANT ON K (THE PERSON CHOSEN)
C           THROUGH THE Z VALUES ARE READ IN FOR EACH PERSON
C
C     THE DIMENSIONS ARE -
C
C     MATRIX  ROW     COL
C
C     ALPHA    -       R
C     Z       NA      NP
C
      INTEGER I, K, R, NA, NP, RC, NAC, NPC
      REAL    ALPHA, Z
C
      DIMENSION   ALPHA(R), Z(NA,NP)
C
C
      UTIL = ( ALPHA(1) + ALPHA(2) * LOG (Z(I,1)) + ALPHA(3) * Z(I,2)
     1        + ALPHA(4) * Z(I,3) + ALPHA(5) * Z(I,4)
     1        + ALPHA(6) * Z(I,5) ) * ( Z(I,1) ** ALPHA(7) )
C
C     UTIL = 0.0
C     DO 20 L = 1, RC
C        UTIL = UTIL + ALPHA(L) * Z(I,L)
C20   CONTINUE
C
      VALV = UTIL
C
      RETURN
C
      END
```

```
C
      REAL FUNCTION PDERV (I,K,L,ALPHA,Z,R,NA,NP,RC,NAC,NPC)
C
C     ROUTINE TO CALCULATE THE TAYLOR SERIES EXPANSION TERMS -
C
C     PDERV(I,K,L)  =  DEL  ( V(ALPHA,Z   ) )   /  DEL  (ALPHA  )
C                                      IK                       L
C
C     I   =  1 - NA, REFERS TO AN ALTERNATIVE (CHOSEN OR NOT)
C     K   =  1 - Q, REFERS TO AN INDIVIDUAL FROM THE SET OF INDIVIDUALS
C     L   =  1 - R, GIVES THE DCM PARAMETER TO DIFFERENTIATE BY
C
C     WHERE
C
C     V( )   =  IS A COND. INDIRECT UTILITY FUNCTION COMPRISING THE
C               PARAMETER VECTOR - ALPHA - AND VARIABLE VECTOR - Z   -
C                                                                  IK
C     ALPHA   =  (ALPHA  , .... , ALPHA )
C                      1               R
C
C     Z     =  (Z    , .... , Z     ), IS AN 'NP' VECTOR
C      IK        IK             IK
C                  1              NP
C
C     THIS ROUTINE MUST BE SUPPLIED BY THE USER FOR A
C     SPECIFIC CHOICE OF COND. INDIRECT UTILITY FUNCTION  - V() -
C     IT IS SUPPLIED WITH THE VECTORS - ALPHA - AND - Z -
C
C     THE DIMENSIONS ARE -
C
C     MATRIX  ROW      COL
C
C     ALPHA    -        R
C     Z       NA       NP
C
      INTEGER I, K, L, R, NA, NP, RC, NAC, NPC
      REAL    ALPHA, Z
C
      DIMENSION   ALPHA(R), Z(NA,NP)
C
C     20-1-86   ..  MAXIMUM EXPENDITURE MODEL
C                   SEE TABLE 7.2
C
      IF (L.EQ.1) THEN
          PDERV = Z(I,1) ** ALPHA(7)
      ELSE IF (L.EQ.2) THEN
          PDERV = LOG (Z(I,1)) * Z(I,1) ** ALPHA(7)
      ELSE IF (L.EQ.3) THEN
          PDERV = Z(I,2) * Z(I,1) ** ALPHA(7)
      ELSE IF (L.EQ.4) THEN
          PDERV = Z(I,3) * Z(I,1) ** ALPHA(7)
      ELSE IF (L.EQ.5) THEN
          PDERV = Z(I,4) * Z(I,1) ** ALPHA(7)
      ELSE IF (L.EQ.6) THEN
          PDERV = Z(I,5) * Z(I,1) ** ALPHA(7)
      ELSE IF (L.EQ.7) THEN
          PDERV = LOG (Z(I,1)) * VALV (I,K,ALPHA,Z,R,NA,NP,RC,NAC,NPC)
      ENDIF
C
C     PDERV = Z(I,L)
C
      RETURN
C
      END
```

```
C
      SUBROUTINE RECODEX (G,Q,P,QC,PC)
C
C     ROUTINE TO RECODE THE CONT. CHOICE MODEL DATA VECTORS -
C
C     X  IS A SUB-MATRIX OF G
C      J
C
C     X     =  (X     , .... , X    ), IS A P VECTOR
C      JK        JK             JK
C                  1               P
C
C     K  =  1 - Q, REFERS TO AN INDIVIDUAL FROM THE SET OF INDIVIDUALS
C                  CHOOSING J
C
C     THIS ROUTINE MUST BE SUPPLIED BY THE USER. IT ALLOWS SIMPLE
C     MANIPULATION OF ELEMENTS OF THE X MATRIX, WHICH BY THIS STAGE HAVE
C     BEEN INCORPORATED INTO THE G MATRIX. FOR INSTANCE, THE X MATRIX
C     INPUTTED INTO THIS PROGRAM MIGHT REPRESENT RAW DATA, BUT IN
C     ESTIMATING THE CONTINUOUS CHOICE MODEL THE USER TOOK A LOG OF
C     ONE OF THE ELEMENTS OF X. THIS ROUTINE ALLOWS SUCH A MANIPULATION
C     TO BE INCORPORATED IN CALCULATING THE CORRECTED VAR/COV MATRIX.
C     THE ROUTINE IS SUPPLIED WITH THE MATRIX - G -
C
C     THE DIMENSIONS ARE -
C
C     MATRIX  ROW     COL
C
C     G       Q       P+1
C     X       Q       P
C
      INTEGER Q, P, QC, PC
      REAL    G
C
      DIMENSION   G(Q,P+1)
C
C
C     20-1-86  ..  MAXIMUM EXPENDITURE MODEL
C                  SEE TABLE 7.2
C
      DO 20 I = 1, QC
         G(I,2) = LOG (G(I,2))
 20   CONTINUE
C
      RETURN
C
      END
```

```
C
      REAL FUNCTION VALE (K,BETA,G,Q,P,QC,PC)
C
C     ROUTINE TO CALCULATE THE PREDICTED VALUES FOR THE CCM DEPENDENT
C     VARIABLE, IGNORING THE SC TERM -
C
C                         T
C     VALE(K)  =  BETA   X
C                          JK
C
C     K  =  1 - Q, REFERS TO AN INDIVIDUAL FROM THE SET OF INDIVIDUALS
C                 CHOOSING ALTERNATIVE J
C
C     WHERE
C
C     THE CONT. CHOICE MODEL IS A LINEAR COMBINATION OF PARAMETER VECTOR
C     - BETA - AND VARIABLE VECTOR - X    - (X   IS A SUB-MATRIX OF G)
C                                   JK      J
C
C     BETA   =   (BETA  , .... , BETA )
C                     1              P
C
C     X     =   (X    , .... , X    ), IS A P VECTOR
C      JK         JK            JK
C                  1              P
C
C     THIS ROUTINE MUST BE SUPPLIED BY THE USER FOR A SPECIFIC
C     FORM OF CONTINUOUS CHOICE MODEL
C     IT IS SUPPLIED WITH THE VECTOR - BETA - AND MATRIX - G -
C
C     THE DIMENSIONS ARE -
C
C     MATRIX  ROW     COL
C
C     BETA     -       P
C     G        Q       P+1
C     X        Q       P
C
      INTEGER K, Q, P, QC, PC
      REAL    BETA, G
C
      DIMENSION   BETA(P), G(Q,P+1)
C
C
C     20-1-86  ..  MAXIMUM EXPENDITURE MODEL
C                  SEE TABLE 7.2
C
      EXPEND = 0.0
      DO 20 L = 1, PC
         EXPEND = EXPEND + BETA(L) * G(K,L)
   20 CONTINUE
      VALE = EXPEND
C
      RETURN
C
      END
```

CHAPTER 8

SUMMARY OF THE MAJOR CONTRIBUTIONS OF THIS THESIS AND SOME SUGGESTIONS ON POSSIBLE FUTURE RESEARCH

The contributions of this study to the understanding of shopping behaviour pervade the previous pages of this thesis. Possibly the most significant advance has been the development of a comprehensive economic theory of shopping destination choice. This development has firmly placed the study of shopping behaviour within the wider context of economic consumption analysis. By so doing it has opened the way for findings from mainstream economic consumer theory to be applied to the study of shopping behaviour.

In this study relationships unearthed in mainstream economic consumer theory have been applied in two principal areas. Firstly, the indirect utility function properties specified by Diewert (1974) were used to check the validity of the conditional indirect utility functions estimated in Chapter 7. Secondly, Roy's identity was used to establish a relationship between shopping destination and expenditure decisions. A by-product of this research was a demonstration that the linear form for the conditional indirect utility functions, as typically used in past work estimating discrete shopping destination choice models, can at best be regarded as a first order approximation to the true non-linear form.

The empirical counterpart to the theoretical link established by application of Roy's identity was an inter-related model of shopping destination and expenditure choices. The structure of the empirical model closely aligned with theoretical considerations. Moreover, the statistical relationship between submodels developed for the destination and expenditure decisions was duly recognised by demarcating the system within the set of sample selectivity models. Although sample selectivity models have been used over a number of years, econometric advances to emerge only within the last two years have permitted the use of these techniques in polychotomous choice situations. To the author's knowledge, this is the first application of the technique developed by Lee (1984) with a non-linear conditional logit model. This study can thus be seen as pioneering in a strictly econometric sense. More generally, this work drives at the heart of ongoing concerns about the economic impacts of retail planning (see Chapter 1) and the effect of transport policies on retail activity (e.g. Atherton and Eder 1982, Kern and Lerman 1982, Loudon and Coogan 1985, Waters 1986). It also represents a significant generalization of possibly the most widely used model in retail activity prediction, namely, the Huff model.

The contributions of the study go further than this. Chaper 5 contained a detailed analysis of choice set variations and the impact these exert on MNL parameter estimates. Although this is a difficult area, it is felt that work reported herein has progressed knowledge by:

(i) demonstrating different parameter values to emerge from MNL models estimated using reported and analyst assigned choice sets and showing that these differences may be inferred from theory,

(ii) constructing the scaffolding of a theory of choice set determination based on economic search theory and random utility theory, and

(iii) showing that reported choice sets from one data set conform to the theory developed.

A final significant contribution involved investigating the linkages between categories of food shopping. Whereas past studies have tended either to lump all shopping together in an homogenous group or to have concentrated on just one category of shopping in isolation from other categories, the model of Chapter 6 traces micro level linkages between food shopping activities. Allowance for these linkages was shown to improve predictions of shopping behaviour.

It remains to bond the multi-trip multi-purpose shopping travel analysis with the discrete/continuous aspects of shopping choice. Two possible mechanisms to do this are covered in the remainder of this thesis.

A fairly obvious way of combining the multi trip multi-purpose shopping travel analysis of Chapter 6 with the discrete/continuous shopping choice analysis of Chapter 7 is to interpret the g_is in Chapter 3 (especially equation 3.4) as consumption of food shopping products (meat, groceries, greengroceries). The price indices may then be specified as $p_i = p_i^* (MPRICE_i, GPRICE_1, VPRICE_i) = a_1(MPRICE_i)^{a_2} \times (GPRICE_i)^{a_3} (VPRICE_i)^{a_4}$ and the t_i and c_i in the time and income constraints as the travel time and cost associated with shopping pattern i. Under the respecification the model system of Chapter 7 would involve the discrete choice of shopping pattern and the continuous choice of food shopping expenditure level. The computational effort of estimating this model would be no more burdensome than that associated with the model of Chapter 7. The only reason for not estimating such a model in the current study was data size restrictions.

Certain disadvantages in simultaneously estimating shopping patterns, however, were noted in Chapter 2. Recent advances in econometrics, in the analysis of panel data, offer the potential for an adequate representation of behaviour using sequential models of travel patterns. A reasonably general representation of the conditional indirect utility function associated with travel choice i at the time t may be given by:

$$V_{iqt} = \overline{V}_{iqt} \, (Z_{iqt}, I_{q(t-1)}, I_{q(t-2)}, \ldots, I_{q1}, V_{iq(t-1)}, V_{iq(t-2)}, \ldots, V_{iq1}) + \epsilon_{iq} + \epsilon_{iqt} \qquad (8.1)$$

where terms are as before with the t subscript referring to time periods.

The two error terms identified in this equation draw attention to three concepts crucial to the correct modelling of choices over time; heterogeneity, state dependence and nonstationarity. Heterogeneity may be defined as the variation which exists between individuals due to observed and unobserved exogenous influences. Observed effects may be empirically picked up through appropriate inclusion of contemporaneous and lagged exogenous variables. The error terms ϵ^{*}_{qi} are specifically designed to capture time invariant unobserved exogenous influences. State dependence is used to refer to the intertemporal relationships in choice behaviour; how current choices are dependent upon past choices and future choices on current choices. Hensher and Wrigley (1986) identify several sources of state dependence, two of which are represented by Markov and semi-Markov processes. Nonstationarity refers to temporal changes in individual choice behaviour arising from variations in the values of observed or unobserved exogenous variables.

Discussion drawn from Heckman (1981) will serve to indicate why the separation of these concepts is so important. Suppose for an event under study there exists an observed dependence between current and past behaviour. There are two possible explanations for this. One is that past experience truly alters behaviour, so that two individuals in identical circumstances in the current period will make different choices if their past experience has not also be identical. This explanation implies that as a consequence of experiencing an event, preferences or constraints relevant to future choices are altered. The other explanation is that unobserved characteristics of individuals which are stable over time affect their propensity to experience the event. This will result in the appearance of current behaviour being dependent on past behaviour.

A simple example from the labour supply literature may used to illustrate this last point. The example is from a study by Lancaster and Nickell (1980) looking at the probability of obtaining work. Assume that there exists an omitted variable termed 'motivation' which is positively related to the chances of obtaining work. Then successive panel samples over time will contain a higher and higher proportion of unemployed individuals deficient in motivation. Consequently correlation may be expected between the past behaviour variable 'duration of unemployment' and the probability of obtaining work. This phenonema is generally termed spurious state dependence. Broadly defined, differences in individual motivations may be categorised as population heterogeneity. Considerations such as these have led Hensher and Wrigley (1984, p.13) to conclude: 'Simply adding a series of lagged exogenous variables does not guarantee accounting for true intertemporal dependence'.

Returning to equation (8.1), the problem is: how can it be simplified to yield a suitable, but manageable, model for empirical estimation? At least two principal approaches are available.

One approach (e.g. Hensher 1984b) is to specify $\overline{V}$ in terms of current period and lagged variables, such that:

$$\overline{V}_{qjt} = Z_{qjt}\,\alpha + \sum_{\ell=1}^{t-1} Z_{qj(t-\ell)}\,\delta^{\ell}\,\alpha \qquad (8.2)$$

where α and δ are parameter vectors. Initial conditions, one contributing factor to unobserved heterogeneity, can be set through appropriate expansion of the error terms. In this approach exogenous, rather than endogenous, variables are lagged to avoid problems of serial correlation.

The second approach involves a more direct consideration of the error terms in equation (8.1). In developing the model assume, initially, stationarity and no state dependence, so that: $\overline{V}_{iqt} = \overline{V}(Z_{iqt}, \alpha)$. Applying standard assumptions concerning the distribution of the ϵ_{iqt}, the choice probabilities can be expressed as:

$$P_{qjt} = \frac{\exp(V_{qjt} + \epsilon^{*}_{qj})}{\sum \exp(V_{qit} + \epsilon^{*}_{qi})} \qquad (8.3)$$

Before the model of equation (8.3) can be applied it is necessary to remove the error terms ϵ^{*}_{qi} . One method is to assume a distributional form for the outcome probabilities P_{qjt} at each level of the variables contained in Z_{qjt} (Heckman and Willis 1977). The most commonly used distribution is the multi variate beta or Dirichlet distribution (Dunn and Wrigley 1985). Recently, Davies (1984) and Davies and Pickles (1984) have generalized the model implied by equation (8.3), albeit in a binary context, to incorporate time varying exogenous variables and state dependence effects.

Although still in its infancy, the research outlined above opens the way for a modelling amalgamation of shopping behaviour with other activity and travel choices to an extent not possible with simple simultaneous models. Contained in this is the challenge to develop truly dynamic models of shopping behaviour, incorporating both discrete and continuous choice components. This is rapidly becoming within reach. Already in another transport related area work is in progress (Hensher 1985) to tackle such a goal.

REFERENCES

ADLER, T.J. and M.E. BEN-AKIVA (1976). Joint choice model for frequency, destination and travel mode for shopping trips. Transportation Research Record, No. 569, pp. 136-150.

______________ (1979) A theoretical and empirical model of trip chaining behaviour. Transportation Research, Vol. 13B, pp. 243-257.

AMEMIYA, T. (1975). Qualitative response models. Annals of Economic and Social Measurement, Vol 4. pp 363-372.

AMPT, E., A.J. RICHARDSON and W. BROG (1985). The role of empirical data for transport policy analysis. Proceedings of the 10th Australian Transport Research Forum, Melbourne, Vol 2, pp. 275-292.

ANSAH, J.A. (1977). Destination choice set definition in travel behaviour modelling. Transportation Research, Vol. 11, pp. 127-140.

ARNOLD, S.J., T.H. OUM and D.J. TIGERT (1983). Determinant attributes in retail patronage : seasonal, temporal, regional, and international comparisons. Journal of Marketing Research, Vol. 20, pp. 149-157.

ATHERTON, T.J. and E.S. EDER (1982). Impacts of CBD fare-free transit on retail sales. Paper presented at the 61st Annual Meeting of the Transportation Research Board, Washington D.C.

AUSTRALIAN BUREAU OF STATISTICS (1972). Travel to work, school and shop in the Adelaide statistical division. Catalogue No. 9201.4, Adelaide.

BARNARD, P.O. (1981). The Adelaide travel demand and time allocation study: questionnaire forms, interview and coding manuals. Australian Road Research Board, AIR 352-2, Vermont, Victoria. 121pp.

______________ (1982). Development of simultaneous mode/destination models for food shopping travel choices. Australian Road Research Board. Internal Report AIR 352-3. Vermont, Victoria.

______________ (1985). Evidence of trip under-reporting in Australian transportation study home interview surveys and its implications for data utilization. In E.S. AMPT, A.J. RICHARDSON and W. BROG (eds), New Survey Methods in Transport, VNU Scientific Press, Amsterdam.

_____________ (1986). Use of an activity diary survey to examine travel and activity reporting in a home interview survey. Transportation, Vol 16, (forthcoming).

BARNARD. P.O. and R.E. BRINDLE (1985). Traffic generation estimation - suggestions on new research directions. Proceedings of the 10th Australian Transport Research Forum, Melbourne, Vol. 1, pp. 43-62.

BARNARD, P.O. and R. NEIL (1985). Choice set specification and the multinomial logit model. Discussion Paper, No. 943, Australian Road Research Board, Vermont South, Victoria.

BARNOW, B.S., G.G. CAIN and A.S. GOLDBERGER (1981). Issues in the analysis of selectivity bias. In W.E. STROMSDORFER and G. FARKAS (eds), Evaluation Studies Review Annual, Vol. 5, pp. 43-59. Beverley Hills: Sage.

BARTON ASCHMAN ASSOCIATES, INC., and CAMBRIDGE SYSTEMATICS, INC. (1976). Task termination report : internal CBD travel demand modelling. Consultant Report, Los Angeles.

BEGGS, S., S. CARDELL and J. HAUSMAN (1981). Assesing the potential demand for electric cars. Journal of Economics, Vol. 16, pp. 1-19.

BEN-AKIVA, M.E. (1977). Choice models with simple choice set generating processes. Working paper. Dept of Civil Engineer, Mass. Institute of Technology, Cambridge, M.A.

BEN-AKIVA, M.E. and S.R. LERMAN (1979). Disaggregate travel and mobility choice models and measures of accessibility. In D.A. HENSHER and P.R. STOPHER (eds), Behavioural Travel Modelling, Croom Helm, London, pp. 654-679.

BEN-AKIVA, M.E. and J. SWAIT (1985). Choice models with simple probabalistic choice set generation processes. Transportation Research B (forthcoming).

BEN-AKIVA, M.E., L. SHERMAN and J. KULLMAN (1979). Non home based models. Transportation Research Record, No. 673, pp. 128-133.

BEN-AKIVA, M.E., S.R. LERMAN, D. DAMM, J. JACOBSON, S. PITSCHKE, G. WEISBROD, and R. WOLFE (1980). Understanding, Prediction and Evaluation of Transportation Related Consumer Behavior. Report prepared for the Office of University Research, Research and Special Programs Administration, U.S. Department of Transportation, Washington.

BERK, R.A. (1983). An introduction to sample selection bias in sociological data. American Sociological Review, Vol. 48, pp. 386-98.

BERKOVIC, J. AND J. RUST (1985). A nested logit model of automobile holdings for one vehicle households. Transportation Research, Vol 19B, pp 275-285.

BLACK, W.C. (1984). Choice-set definition in patronage modelling. Journal of Retailing, Vol. 60, pp. 63-85.

BOCK, R.D. and L.V. JONES (1968). The Measurement and Prediction of Judgement and Choice. San Francisco: Holden Day.

BOERSCH-SUPAN, A. (1985). Hierarchical choice models and efficient sampling: with applications in the demand for housing. Methods of Operations Research, Vol. 50, pp. 175-186.

BROG, W. and E. ERL (1981). Application of a model of individual behaviour (situational approach) to explain household activity patterns in an urban area and to forecast behavioural changes. Paper presented to International Conference on Travel Demand Analysis: Activity-based and Other New Approaches, Oxford University July, 1981.

BROWN, L.A. and E.G. MOORE (1970). The intra-urban migration process : a perspective. Geografiska Annaler, Vol. 52B, pp. 1-13.

BROWN, L.A., E.J. MALECKI and S.G. PHILLIBER (1977). Awareness space characteristics in a migration context. Environment and Behaviour, Vol. 9, pp. 335-348.

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

BURNETT, P. (1973). The dimensions of alternatives in spatial choice processes. Geographical Analysis, Vol. 5. pp. 181-204.

__________ (1974). Disaggregate behavioural models of travel decisions other than mode choice ; a review and contribution to spatial choice theory. Transportation Research Board Special Report, No. 149, Washington D.C., pp. 207-222.

BURNETT, P. and S. HANSON (1982). The analysis of travel as an example of complex human behaviour in spatially-constrained situations : definition and measurement issues. Transportation Research, Vol. 16A, pp. 87-102.

CHARLES RIVER ASSOCIATES, INC. (1976). Disaggregate Travel Demand Models Vol. II. Project 8-13, Phase 1 Report, National Highway Research Program, Charles River Associates, Boston.

COCHRANE, R.A. (1975). A possible economic basis for the gravity model. Journal of Transport Economics and Policy, Vol. 9, pp. 34-49.

CHRISTENSEN, L.R., D.W. JORGENSON and L.J. LAU (1975). Transcendental logarithmic utility functions. American Economic Review, Vol. 65, pp. 217-235.

CLARKE, M.I., M.C. DIX and P.M. JONES (1985). Household Activity - Travel Patterns in Adelaide. Interim Report to the Director-General of Transport South Australia, Oxford Transport Studies Unit, Oxford, 149pp.

CURHAN, R.C. (1972). The relationship between shelf space and unit sales in supermarkets. Journal of Marketing Research, Vol. 9, pp. 406-412.

DALY, A. (1978). Issues in the estimation of journey attribute values. In D.A. HENSHER and M.Q. DALVI (eds.), Determinants of Travel Choice, Saxon House, Teakfield, Farnborough, England, pp. 335-357.

_____________ (1982). Estimating choice models containing attraction variables. Transportation Research, Vol. 16B, pp. 5-16.

DAMM, D. (1982). Parameters of activity behaviour for use in travel analysis. Transportation Research, Vol. 16A, pp. 135-148.

_____________ (1984). The integration of activity and transportation analysis for use in public decision making. Transport Policy and Decision Making, Vol. 2, pp. 249-269.

DAMM, D. and S.R. LERMAN (1981). A theory of activity scheduling behaviour. Environment and Planning A, Vol. 13, pp. 703-718.

DAVID, H.A. (1970). Order Statistics. Wiley, New York. 272 pp.

DAVIES, R.B. and R. CROUCHLEY (1985). Controlling for omitted variables in the analysis of panel and other longitudinal data. Geographical Analysis, Vol. 17, pp. 1-15.

DAVIES, R.B. and A.R. PICKLES (1985). A joint trip timing destination choice model for grocery shopping, including inventory effects and parametric control for omitted variables. Working paper submitted for publication to Transportation Research B.

DEATON, A.S. and J. MUELLBAUER (1980a). Economics and Consumer Behaviour. Cambridge University Press, Cambridge 450pp.

_____________ (1980b). An almost ideal demand system. *American Economic Review*, Vol. 70, pp. 312-336.

DEVINE, D.G. and B.W. MARION (1979). The influence of consumer price information on retail pricing and consumer behaviour. *American Journal of Agricultural Economics*, Vol 61, pp 228-237.

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.

DE SERPA, A.C. (1971). A theory of the economics of time. *Economic Journal*, Vol. 81, pp. 828-846.

DOMENCICH, T.A. and D. McFADDEN (1975). *Urban Travel Demand*. North Holland/Americal Elsevier, New York.

DUAN, N., W.G. MANNING, C.N. MORRIS and J.P. NEWHOUSE (1983). A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics*, Vol. 1, pp. 115-26.

_____________ (1984). Choosing between the sample-selection model and multi-part model. *Journal of Business and Economic Statistics*, Vol. 2, pp. 283-89.

DUBIN, J. and D. McFADDEN (1980). An econometric analysis of residential electrical appliance holdings and consumption. Working Paper, Department of Economics, Massachusetts Institute of Technology.

_____________ (1984). An econometric analysis of residential electrical appliance holdings and consumption. *Econometrica*, Vol. 52, pp. 345-362.

DUNN, R. and N. WRIGLEY (1985). Beta-logistic models of urban shopping centre choice. *Geographical Analysis*, Vol. 17, pp. 95-113.

EAGLE, T.C. (1984). Parameter stability in disaggregate retail choice models - experimental evidence. *Journal of Retailing*, Vol. 60, pp. 101-123.

FLEET, C.R. and S.R. ROBERTSON (1968) Trip generation in the transportation planning process. *Highway Research Record*, No. 240, pp. 13-25.

ELLIOTT, D. and J. HOLLENHORST (1981). Sequential unordered logit applied to college selection with imperfect information. *Behavioural Science*, Vol. 26, pp 366-378.

FISHER, F.M. and K. SHELL (1971). Taste and quality change in the pure theory of the true cost of living index. In Z. GRILICHES (ed), Price Indexes and Quality Change : Studies in New Methods of Measurement, Harvard University Press, Cambridge, MA.

FLEET, C.R. and S.R. ROBERTSON (1968). Trip generation in the transportation planning process. Highway Research Record, No. 240, pp. 101-123.

GAUDRY, M. and M. DAGENAIS (1979). The dogit model. Transportation Research, Vol. 13B, pp. 105-111.

GAUTSCHI, D.A. (1981). Specification of patronage models for retail centre choice. Journal of Marketing Research, Vol. XVIII, pp. 162-174.

GHOSH, A. (1984). Parameter nonstationarity in retail choice models. Journal of Business Research, Vol. 12, pp. 425-436.

GRAHAM, N.R. and K.W. OGDEN (1979). A study of travel linkages : implications for urban transport planning. Proceedings of the 5th Australian Transport Research Forum, pp. 441-457.

GUNN, H.F., M.E. BEN-AKIVA and M.A. BRADLEY (1986). Tests of the scaling approach to transferring disaggregate demand moels. Transportation Research Record (forthcoming).

GUY, C.M. (1984). The estimation of retail turnover for planning purposes. The Planner, Vol. 70 No. 5, pp. 12-14.

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

HANSON, S. (1979). Urban travel linkages : a review. In D.A. HENSHER and P.R. STOPHER (eds), Behavioural Travel Modelling, Croom Helm, London, pp. 81-100.

__________ (1980). The importance of the multi-purpose journey to work in urban travel behaviour. Transportation, Vol. 9, pp. 229-248.

HAUSMAN, J.A. (1981). Exact consumer's surplus and deadweight loss. American Economic Review, Vol. 71, pp. 662-679.

HAY, J.W. (1980). Occupational Choice and Occupational Earnings : Selectivity Bias in a Simultaneous Logit-OLS Model. PhD Dissertation, Yale University.

HAY, J.W. and R.J. OLSEN (1984). Let them eat cake: a note on comparing alternative models of the demand for medical care. Journal of Business and Economic Statistics, Vol 2, pp. 279-82.

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, Vol. 5, pp. 475-592.

______________ (1979). Sample selection bias as a specification error. Econometrica, Vol. 46, pp. 931-959.

______________ (1981). Statistical models for discrete panel data. In C.F. MANSKI and D. McFADDEN (eds), Structural Analysis of Discrete Data : with Econometric Applications, M.I.T. Press, Cambridge, pp. 114-178.

HECKMAN, J. and R. WILLIS (1977). A beta-logistic model for the analysis of sequential labour force participation by married women. Journal of Political Economy, Vol. 85, pp. 27-58.

HEGGIE, I.G. (1977). Putting behaviour into behavioural models of travel demand. Working Paper No. 22, Oxford University, Transport Studies Unit, Oxford, England.

HEGGIE. I.G. and P.M. JONES (1978). Defining domains for models of travel demand. Transportation, Vol. 7, pp. 119-135.

HEMMENS, G.C. (1966). The Structure of Urban Activity Linkages. Urban Studies Monograph, Center for Urban and Regional Studies, University of North Carolina, Chapel Hill.

______________ (1970). Analysis and simulation of urban activity patterns. Socio Economic Planning Science, Vol. 4, pp. 53-66.

HENSHER, D.A. (1976). The structure of journeys and nature of travel patterns. Environment and Planning, Vol. 8, pp. 665-672.

______________ (1979). Individual choice modelling with discrete commodities : theory and application to the Tasman Bridge reopening. Economic Record, Vol. 55, pp. 243-260.

______________ (1981). A practical concern about the relevance of alternative-specific constants for new alternatives in simple logit models. Transportation Research, Vol. 15B, pp. 407-410.

______________ (1984a). Accounting for curvilinearity in explanatory variables: a simplified alternative. International Journal of Transport Economics, Vol. 11, pp. 43-51.

______________ (1984b). Model specification for a quasi-dynamic discrete-continuous choice automobile demand system in discrete time using panel data. Dimensions of Automobile Demand Project, Working Paper No. 11, 75pp.

______________ (1986). Full information maximum likelihood estimation of a nested logit mode choice model. Review of Economics and Statistics, Vol. 68, (forthcoming).

HENSHER, D.A. and N. WRIGLEY (1984). Statistical modelling of discrete choices with panel data. Dimensions of Automobile Demand Project, Working Paper No. 16, 70pp.

______________ (1986). Statistical modelling of discrete choices in discrete time with panel changes. In MINISTRY OF TRANSPORT AND PUBLIC WORKS (eds), Behavioural Research for Transport Policy, VNU Science Press, The Netherlands, 494pp.

HICKS, J.R. (1936). Value and Capital. Oxford University Press, Oxford.

HILDEBRAND, F.B. (1956). Introduction to Numerical Analysis. McGraw-Hill, New York.

HILL, M.A. (1983). Female labour force participation in developing and developed countries - consideration of the informal sector. Review of Economics and Statistics, Vol. 65, pp. 459-468.

HOROWITZ, J. (1980). A utility maximising model of the demand for multi-destination non-work travel. Transportation Research, Vol. 14B, pp. 369-386.

HORTON, F.E. and P.W. SHULDINER (1967). The analysis of land use linkages. Highway Research Record, No. 165, pp. 96-107.

HORTON, F.E. and W.E. WAGNER (1968). A Markovian analysis of urban travel behaviour : pattern response by socio-economic-occupational groups. Highway Research Record, No. 283, pp. 19-29.

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

HUMMON, N.P. and L.D. BURNS (1981). Automobile tours : how we use our cars. GMR-3771, General Motors Research Laboratories. Warren, Michigan, 38pp.

JOHNSON, N.L. and S. KOTZ (1970). Distributions in Statistics: Continuous Univariate Distributions. Houghton Mifflin, Boston.

______________ (1972). Distributions in Statistics: Continuous Multivariate Distributions. Wiley, New York.

JONES, P.M. (1976). The analysis of modelling of multi-trip and multi-purpose journeys. Oxford University Transport Studies Unit, Working Paper No. 6.

_______________ (1978). Identifying the determinants of destination choice. In D.A. HENSHER and M.Q. DALVI (eds), Determinants of Travel Choice, Teakfield, Farnborough, England.

JONES, P.M., M.C. DIX, M.I. CLARKE, and I.G.HEGGIE (1980). Understanding Travel Behaviour. Final Report to the SSRC, Oxford University, Transport Studies Unit, Ref 119/PR.

JONES, P.J., M.C. DIX, M.I. CLARK and I.G. HEGGIE (1983). Understanding Travel Behaviour. Gower, Hampshire, England.

KENNY, L.W., L.F. LEE, F.S. MADDALA and P. TROST (1979). Returns to college education: an investigation of self-selection bias based on the project talent data. International Economic Review, Vol. 20, pp. 775-789.

KERN, C.R. and S.R. LERMAN (1982). Models for predicting the impact of transportation policies on retail activity. Paper presented at the 61st Annual Meeting of the Transportation Research Board, Washington D.C.

KITAMURA, R. (1983). Sequential, history-dependent approach to trip-chaining behaviour. Transportation Research Record, No. 944, pp. 13-22.

_______________ (1984). Incorporating trip chaining into analysis of destination choice. Transportation Research, Vol. 18B, pp. 67-81.

KITAMURA, R. and M. KERMANSHAH (1983). Identifying time and history dependencies of activity choice. Transportation Research Record, No. 944, pp. 22-30.

_______________ (1984). A sequential model of interdependent activity and destination choices. Paper presented at the 63rd Annual Meeting of the Transportation Research Board, Washington, D.C.

KITAMURA, R., L.P. KOSTYNUIK and K. TING (1979). Aggregation in spatial choice models. Transportation Science, Vol. 13, pp. 325-342.

KITAMURA, R., L.P. KOSTYNIUK and M.J. UYENO (1981). Basic properties of urban time-space pathes : empirical tests. Transportation Research Record, No. 794, pp. 8-19.

KITAMURA, R. and T.N. LAM (1983). A time dependent Markov renewal model of trip chaining. In V.F. HURDLE, E. HAUER and G.N. STEWART (eds), Proceedings of the Eighth International Symposium of Transportation and Traffic Theory, Toronto University Press, Toronto, Canada, pp. 376-402.

KMENTA, J. (1971). Elements of Econometrics. Collier Macmillan, London, 655pp.

KOPPELMAN, F.S. (1980). Consumer analysis of travel choice behaviour. Journal of Advanced Transportation, Vol. 14, pp. 133-161.

KOPPELMAN, F.S. and HAUSER, J.R. (1977). Consumer travel choice behaviour : an empirical analysis of destination choice for non-grocery shopping trips. Transportation Center Research Report 414-09, Northwestern University, Evanston, IL.

______________ (1978). Destination choice behaviour for non-grocery-shopping trips. Transportation Research Record, No. 673, pp. 157-165.

KOSTYNUIK, L.P. (1975). A Behavioural Choice Model for the Urban Shopping Activity. Ph.D. Dissertation, Department of Civil Engineering, State University of New York at Buffalo.

LANCASTER, A. and S. NICKELL (1980). The analysis of re-employment probabilities for the unemployed. Journal of Royal Statistical Society A, Vol. 143, Part 2, pp. 141-165.

LANDAU, U., J.N. PASHKER and B. ALPERN (1982). Evaluation of activity constrained choice sets to shopping destination choice modelling. Transportation Research, Vol. 16A, pp. 199-207.

LAU, L.J. (1978). Testing and imposing monotonicity, convexity and quasiconvexity constraints. In M. FUSS and D. McFADDEN (eds), Production Economics : A Dual Approach to Theory and Applications, North Holland, Amsterdam, pp. 409-469.

LEE, L.F. (1978). Unionism and wage rates: a simultaneous equation model with qualitative and limited dependent variables. International Economic Review, Vol. 19, pp. 415-433.

______________ (1983). Generalised econometric models with selectivity. Econometrica, Vol. 51, pp. 507-512.

______________ (1982). Some approaches to the correction of selectivity bias. Review of Economic Studies, Vol. 49, pp. 355-372.

LEE, T.R. (1954). A Study of Urban Neighbourhood. Ph.D. dissertation, University of Cambridge, Cambridge, England.

LERMAN, S.R. (1979). The use of disaggregate choice models in semi-Markov process models of trip chaining behaviour. Transportation Science, Vol. 13, pp. 273-291.

LEVIN, I.P., J.J. LOUVIERE, R.J. MEYER and D.H. HENLEY (1979). Perceived versus actual modal travel times and costs for the work trip. Institute of Urban and Regional Research, Technical Report 120, The University of Iowa, Iowa.

LOUDON, W.R. and M.A. COOGAN (1985). The work-based retail activity model : a tool for downtown development planning. Paper presented at the 64th Meeting of the Transportation Research Board, Washington D.C.

LOUVIERE, J.J. and R.J. MEYER (1981). A composite attitude - behaviour model of traveller decision making. Transportation Research, Vol. 15B, pp. 411-420.

LOUVIERE, J.J., E.M. WILSON and M. PICCOLO (1979). Application of psychological measurement and modelling to behavioural travel-demand analysis. In D.A. HENSHER and P.R. STOPHER (eds.), Behavioural Travel Modelling, Croom Helm, London, pp. 713-738.

MADDALA. G.S. (1983). Limited-Dependent and Qualitative Variables in Econometrics. Econometric Society Publication No. 3, Cambridge University Press, Cambridge.

MALHOTRA, N.K. (1984). The use of linear logit models in marketing research. Journal of Marketing Research, Vol. XXI, pp. 20-31.

MANNERING, F.L. (1985). Selectivity bias in models of discrete / continuous choice: an empirical analysis. Paper presented at the Annual Meeting of the Transportation Research Board, 20pp.

MANNERING, F.L. and D.A. HENSHER (1986). Discrete / continuous econometric models and their application to transport analysis. Transportation Research, Vol. 20B, (forthcoming).

MANNERING, F.L. and C. WINSTON (1983). Dynamic models of household vehicle ownership and utilization : an empirical analysis. Centre for Transportation Studies, Massachusetts Institute of Technology.

MANSKI, C.F. (1977). The structure of random utility models. Theory and Decision, Vol. 8, pp. 229-254.

MARBLE, D.F. (1964). A simple Markovian model of trip structures in a metropolitan region. Papers of the Regional Science Association, Western Section, pp. 150-156.

MARSCHAK, J. (1959). Binary choice constraints and random utility indicators. In K.J. ARROW (ed), Mathematical Methods in the Social Sciences, Standford University Press, Standford.

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

McFADDEN, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. ZAREMBKA (ed), Frontiers in Econometrics, Academic Press, New York, pp. 105-142.

______________ (1975). The revealed preferences of a government bureaucracy. Bell Journal of Economics, Vol 7, pp. 55-72.

______________ (1978). Modelling the choice of residential location. In A. KARLQVIST, L. SNICKARS and J.W. WEILBULL (eds), Spatial Interaction Theory and Planning Models, North Holland, Amsterdam.

______________ (1979). Quantitative methods for analysing travel behaviour of individuals: some recent developments. In D.A. HENSHER and P.R. STOPHER (eds), Behavioural Travel Modelling, Croom Helm, London, pp. 279-318.

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

McFADDEN, D. and F. REID (1975). Aggregate travel demand forecasting from disaggregate behavioural models. Transportation Research Record, No. 534, pp. 24-37.

McFADDEN, D., C. WINSTON and A. BOERSCH-SUPAN (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.

MEYER R. (1980). Theory of destination choice set formation under informational constraint. Transportation Research Record, No. 750, pp. 6-12.

MICHAELS, R.M. (1974). Behavioural measurement: an approach to predicting travel demand. Transportation Research Board, Special Report No. 149, pp. 55-66.

MILLER, E.J. and M.E. O'KELLY (1982). Estimating shopping destination choice models from travel diary data. Paper presented at the 61st Annual Meeting of the Transportation Research Board, Washington D.C.

MILLER, G.A. (1956). The magic number seven plus or minus two: some limits on our capacity for processing information. Psychological Review, Vol. 63, pp. 81-97.

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

MUELLBAUER, J. (1974). Household production theory, quality, and the 'hedonic technique'. American Economic Review, Vol. 64, pp. 977-994.

MUTHEN, B. and K.G. JORESKOG (1983). Selectivity problems in quasi-experimental studies. Evaluation Review, Vol. 7, pp. 139-174.

NAKANISKI, M. and L.F. COOPER (1974). Parameter estimation for the multiplicative interactive choice model - least squares approach. Journal of Marketing Research, Vol. XI, pp. 303-311.

NATIONAL OPINION RESEARCH CENTRE (1974). Set of memeographed summaries of nationwide polls taken during energy shortages, October 1973 through March 1974.

O'FARRELL, P.N. and J. MARKHAM (1974). Commuter perceptions of public transport work journeys. Environment and Planning A, Vol. 6, pp. 79-100.

OLSEN, R.J. (1980). A least squares correction for selectivity bias. Econometrica, Vol. 48, pp. 1851-1820.

ORTUZAR, J. de D (1982). Fundamentals of discrete multi-modal choice modelling. Transport Reviews, Vol. 2, pp. 47-78.

ORTUZAR, J de D, and P.C.F. DONOSO (1983). The effects of measurement techniques, variable definition and model specification on disaggregate demand model functions. Proceedings of the 11th PTRC Summer Annual Meeting, Transportation Planning Methods Seminar, Vol. P243, pp. 299-310.

OSTER, C.V. (1978). Household tripmaking to multiple destinations : the overlooked urban travel pattern. Traffic Quarterly, Vol. 32, pp. 511-530.

PAK POY, P.G. AND ASSOCIATES (1978). Metropolitan Adelaide Data Base Study Phases 1-5. Reports prepared for the South Australian Department of Transport and Highways Department, Adelaide, Australia.

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

POURIER, D.J. and P.A. RUDD (1981). On the appropriateness of endogenous switching. Journal of Econometrics, Vol. 16, pp. 249-56.

PROJECT BUREAU INTEGRAL TRAFFIC AND TRANSPORTATION STUDIES (1977). The SIGMO Study. Four reports to the Ministry of Transport of the Netherlands, The Hague.

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

REILLY, W.J. (1929). Methods for the Study of Retail Relationships. Bureau of Business Research Studies in Marketing, Publication No. 4, University of Texas, Austin.

RICHARDS, M.G. (1979). A disaggregate shopping model : some retrospective thoughts. In G.R.M. JANSEN, P.H.L.BOVY, J.P.J.M. VAN EST and F. LE CLERCQ (eds), New Developments in Modelling Travel Demand and Urban Systems, Saxon House, England, pp. 38-51.

RICHARDS, M.G. and M.E. BEN-AKIVA (1975). A disaggregate travel demand model. Saxon House, England, 163pp.

ROBBINS, L. (1930). On the elasticity of demand for income in terms of effort. Economica, pp. 123-129.

ROTHSCHILD, M. (1974). Searching for the lowest price when the distribution of prices is unknown. Journal of Political Economy, Vol. 82, pp. 689-711.

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

RUITER, E.R. and M.E. BEN-AKIVA (1978). Disaggregate travel demand models for the San Fransisco Bay Area : system structure, component models and application procedures. Transportation Research Record, No. 673, pp. 121-128.

RUSHTON, G., R.G. GOLLEDGE and W.A.V. CLARK (1967). Formulation and test of a normative model for the spatial allocation of grocery expenditures by a dispersed population. Annals of the Association of American Geographers, Vo. 57, pp. 389-400.

SALMON, I. and M.E. BEN-AKIVA (1983). The use of the life-style concept in travel demand models. Environment and Planning A, Vol. 15, pp. 623-638.

SHEFFI, Y. (1979). Estimating choice probabilities among nested alternatives. Tranportation Research, Vol. 13B, pp. 189-205.

SMALL, K.A. and H.S. ROSEN (1981). Applied welfare economics with discrete choice models. Econometrica, Vol. 49, pp. 105-130.

SOBEL, K.L. (1980). Travel demand forecasting by using the nested multinomial logit model. Transportation Research Record, No. 775, pp. 48-55.

SOMMERVILLE, R. and D. WILMOTH (1985). Towards a retail policy for metropolitan Sydney. In R. STIMSON and R. SANDERSON (eds), Assessing the Economic Impact of Retail Centres, Australian Institute of Urban Studies, AIUS Publication No. 122, pp. 77-95.

SOUTHWORTH, F. (1981). Calibration of multinomial logit models of mode and destination choice. Transportation Research, Vol. 15A, pp. 315-325.

SPARMANN, J.M. and C.F. DAGANZO (1982). Tromp user's manual. University of California, Institute of Transportation Studies, Research Report UCB-ITS-RR-82-4, 84pp.

STIGLER, G.J. (1961). The economics of information. Journal of Political Economy, Vol. 69, pp 213-225.

STOPHER, P.R. (1979). Preference models and destination choice. Proceedings of the 7th PTRC Summer Annual Meeting, Transportation Models Seminar, Vol. P183, pp. 319-334.

STOPHER, P.R. and A.H. MEYBURG (1975). Urban Transportation Modelling and Planning. Lexington Books, D.C. Heath and Co., Lexington, Mass.

SWAIT, J.D. and M.E. BEN-AKIVA (1985). Incorporating random constraints in discrete choice models : an application to mode choice in Sao Paulo, Brazil. Working paper submitted for publication to Transportation Research B.

SWAIT, J.D. Jr., V.J. KOZEL, R.C. BARROS and M.E. BEN-AKIVA (1984). A model system of individual travel behaviour for a Brazilian city. Transport Policy and Decision Making, Vol. 2, pp. 451-480.

TALVITIE, A. and D. KIRSHNER (1978). Specification, transferability and the effect of data outliers in modelling the choice of mode in urban travel. Transportation, Vol. 7, pp. 311-331.

TARDIFF, T.J. (1979). Definition of alternatives and representation of dynamic behaviour in spatial choice models. Transportation Research Record, No. 723, pp. 25-30.

TERZA, J. (1985). Econometric models with normal polytomous selectivity. Economic Letters, Vol. 19, pp. 165-170.

TIMMERMANS, H., R. VAN DER HEIJDEN and H. WESTERVELD (1982). The identification of factors influencing destination choice: an application of the repertory grid methodology. Transportation, Vol. 11, pp. 189-203.

______________ (1984). Decision making between multiattribute choice alternatives : a model of spatial shopping behaviour using conjoint measurements. Environment and Planning A, Vol. 16, pp. 377-387.

TRAIN, K.E. and M. LOHRER (1983). Vehicle ownership and usage : an integrated system of disaggregate demand models, paper presented at the Transportation Research Board Annual Meeting, Washington, D.C.

TRAIN, K.E. and D. McFADDEN (1978). The goods/leisure trade-off and disaggregate work trip mode choice models. Transportation Research, Vol. 12, pp. 349-353.

TROST, R.P. and L.F. LEE (1984). Technical training and earnings : a polychotomous choice model with selectivity. Review of Economics and Statistics, Vol. 66, pp. 151-156.

TRUONG, T.P. and D.A. HENSHER (1985). Measurement of travel time values and opportunity cost from a discrete choice model. Economic Journal, Vol. 95, (forthcoming).

VAN DER HEIJDEN, R. and H. TIMMERMANS (1984). Modelling choice set generating processes via stepwise logit regression procedures : some empirical results. Environment and Planning A, Vol. 16, pp. 1249-1255.

VAN DER HOORN, A. and J. VOGELAAR (1978). SIGMO : disaggregate models for the Amsterdam conurbation. Proceedings of the 6th PTRC Summer Annual Meeting, Transportation Models Seminar, Vol. P168, pp. 87-104.

VARIAN, H.R. (1984). Microeconomic Analysis. New York: W.W. Norton and Company.

VERSTER, A.C.P., I.L. VAN LEEUWEN and M. DE LANGEN (1979). The effects of the location of a peripheral self-service store. In G.R.M. JANSEN, P.H.L. BOVY, J.P.J.M. VAN EST and F. LE CLERCQ (eds), New Developments in Modelling Travel Demand and Urban Systems, Saxon House, England, pp. 52-74.

VICKERMAN, R.W. and T.A. BARMBY (1984). The structure of shopping travel : some developments of the trip generation model. Journal of Transport Economics and Policy, Vol. 18, pp. 109-121.

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

WATANATADA, T. and M.E. BEN-AKIVA (1978). Spatial aggregation of discrete choice models : areawide urban travel sketch-planning model. Transportation Research Record, No. 673, pp. 93-99.

WATERS, D.C. (1986). Use of ensemble averaging and differential comparison to detect retail sales increases by the Bay Area Rapid Transit System. Transportation Research, Vol. 20B, (forthcoming).

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

WHEELER, J.O. (1972). Trip purposes and urban activities linkages. Annals of the American Association of Geographers, Vol. 62, pp. 641-654.

WIGAN, M.R. (1982). Bicycle ownership, use and exposure in Melbourne 1978-79. Australian Road Research Board Internal Report, AIR380-3, Vermont, Victoria 65pp.

WILLIAMS, H.C.W.L. and J.D. ORTUZAR (1979). Behavioural travel theories, model specification and the response error problem. Proceedings of PTRC Summer Annual Meeting, University of Warwick, July 1979.

_____________ (1982). Behavioural theories of dispersion and the mis-specification of travel demand models. Transportation Research, Vol. 16B, pp. 167-219.

WINSHIP, C. and R.D. MARE (1984). Regression models with ordinal variables. American Sociological Review, Vol. 49, pp. 512-525.