



**AgEcon** SEARCH  
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search  
<http://ageconsearch.umn.edu>  
[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*



The Retail Food Industry Center  
Advancing Knowledge About  
Processing, Distribution,  
Sales, and Food Service

Working Paper 96-01  
The Retail Food Industry Center  
Printed Copy \$22.50

**Estimation of Household Brand-size Choice Models  
For Spaghetti Products with Scanner Data**

Changwon Park and Ben Senauer\*

Department of Applied Economics

University of Minnesota

March 1996

\*Changwon Park is a Lecturer in the Department of Economics at Sung Kyun Kwan University in Seoul, Korea, and completed this work as a Ph.D. student at the University of Minnesota. Ben Senauer is a Professor and Director of the Center for International Food and Agricultural Policy in the Department of Applied Economics at the University of Minnesota. They wish to thank Nielsen Marketing Research for access to the household panel scanner data used in this research.

The Retail Food Industry Center is an Alfred P. Sloan Foundation Industry Studies Center.

Working Paper 96-1  
The Retail Food Industry Center

**Estimation of Household Brand-size Choice Models**  
**For Spaghetti Products with Scanner Data**

Changwon Park and Ben Senauer\*

Copyright © 1996 by Park and Senauer. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

The analyses and views reported in this paper are those of the authors. They are not necessarily endorsed by the Department of Applied Economics or by the University of Minnesota.

The University of Minnesota is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation.

Information on other titles in this series may be obtained from The Retail Food Industry Center, University of Minnesota, Department of Applied Economics, 1994 Buford Avenue, 317 COB, St. Paul, MN 55108-6040, U.S.A.

## Estimation of Household Brand-Size Choice Models for Spaghetti Products with Scanner Data

### ABSTRACT

The A.C. Nielson household scanner panel is used to analyze factors affecting brand-size choices for spaghetti. These data link product purchases, which are scanned, with household demographics and market information on the store where purchased. A multinomial logit model and a nested logit model are specified and estimated. In addition to the effects of inventory and brand loyalty, the most interesting results relate to the price elasticities of choice probabilities. These elasticities are much more elastic than those found in traditional food demand analysis. Different brands and sizes of spaghetti products on the same supermarket shelf are very close substitutes for the consumer.

## Estimation of Household Brand-Size Choice Models for Spaghetti Products with Scanner Data

### INTRODUCTION

One of the distinct American food consumption trends of the last decade is the increased intake of pasta. Per capita consumption of durum wheat flour, mainly used in pasta, more than doubled between 1982 and 1993, from 6.1 to 13.5 pounds per person (Putnam and Allshouse, 1994, pp. 18, 58). Pasta, known as a low-fat, inexpensive, and filling food, is often considered a light alternative to more traditional dishes (Bunch and Wendland, 1986, p. 3). Understanding how prices and marketing strategies affect consumer choice behavior for a specific pasta product assists industry planners and retailers in determining optimal price and promotion policies. In addition, knowledge of how choices vary by household characteristics is relevant to predicting consumption trends. Therefore, the purpose of this study is to develop and estimate household choice models for spaghetti products.

The introduction of computerized scanner-checkout systems in U.S. supermarkets in the mid 1970s provided a potential source of new data for economics and market research. Scanner data are very accurate because they directly record the sales activity for each product level with a Universal Product Code (UPC), the bar code on the package. Scanner data provide a view of the competitive environment in which consumers make decisions. They not only tell what consumers buy at which price, but can also identify the other products, prices, and marketing activities such as coupons, redemptions, retail advertising, and shelf-space allocation at the time of purchase.

The major weakness of scanner data has been that no information is provided on the

consumer who made the purchase. However, data are now available which link purchases with consumers through frequent shopper programs or household panels. In the former, stores enroll shoppers in a program which provides them with an identification card that is shown to the cashier during check-out to receive certain benefits, such as a discount. A separate survey is used to collect household information from the participating consumers. In addition, some market research companies have established their own household panels. The panel households are provided with scanners which they use to scan their purchases after returning from shopping. The data used in this research are from the A.C. Nielsen household panel, which links data on product purchases, household demographics, and store information.

Most traditional demand analyses have been conducted with either time-series or cross-sectional data with only a limited degree of product disaggregation. Many demand studies have, for example, used food categories, which include beef, pork, and poultry. Although beef, pork, and poultry are substitutes, they are still much more aggregate than the level of disaggregation provided by scanner data. The brand-size choice alternatives for spaghetti products in this study are much closer substitutes than those in traditional demand analysis.

A number of previous studies in the marketing and agricultural economics areas have been conducted to analyze consumer choice behavior with scanner data. Marketing studies have mainly tried to deal with the consumer's choice response for branded goods to market strategies (Guadagni and Little, 1983; Wagner and Taudes, 1986; Fader and McAlister, 1990; Bucklin and Lattin, 1991; Russell and Kamakura, 1994). The agricultural economics studies have focused on the structure of consumer demand for disaggregated products such as ground beef, roast beef, and steak (Capps, 1989; Capps and Nayga, 1991; Capps and Lambregts, 1991). These research

studies have contributed to our knowledge and understanding of consumer behavior.

However, these previous studies have some common drawbacks. Local or regional data have been used, so that the results do not represent the national consumption pattern. Most research has tried to explain the overall consumers' response to prices or advertising activity at the store level rather than at the consumer level, since the available scanner data were not linked to households. Therefore, few studies have related consumer choice behavior to relevant demographic characteristics. In addition, the national consumer panel data used in this study allows us to analyze responsiveness over time.

In the case where consumers face very close substitutes, like different spaghetti brand-size products on the same supermarket shelf, the demands are not continuous. Choosing a particular product will likely result in zero expenditures on alternative spaghetti products during a particular shopping trip. The changes in demand are not smooth and many products are not purchased by a particular household. Each good is no longer chosen to buy more or less, but is chosen to buy or not to buy, which is in turn, more qualitative than quantitative. Therefore, it is appropriate that demand analysis be approached in a qualitative manner rather than a quantitative one. In this context, appropriate econometric models are used to explain the household's food choice behavior as affected by demographic factors as well as shopping environment factors.

## THEORETICAL FRAMEWORK

This study uses a standard random utility model as its theoretical basis (for examples see McFadden, 1981 or Hanemann, 1984). The consumer is assumed to face a choice decision among products that are substitutes for one another. The consumer's utility function is composed of deterministic attributes that contain some components which are unobservable to the researcher and are treated as random variables. The realized choice decision is assumed to be generated from the consumer's utility maximization process.

### Multinomial Logit Model

We propose a household brand-size choice model for spaghetti products that describes the household's choice probability, given the product category is a linear function of the brand-size attributes and household demographic factors.<sup>1</sup> The household is assumed to treat alternatives independently and to choose one of the choice alternatives which represents the most preferred alternative at the time of choice.

Each household  $h$  ( $h=1, \dots, N$ ) has a choice set  $J$  consisting of alternative choices ( $m=0, 1, \dots, M$ ), where  $m=0$  is the nonpurchase alternative, and  $m=1, \dots, M$  are the brand and package size alternatives for spaghetti, at  $t$  time period ( $t=1, \dots, T$ ). Assume that the indirect utility function  $U^h(t)$  for each household consists of two parts: a non-stochastic term  $V^h(t)$  and a stochastic term (or random error)  $\varepsilon^h(t)$ . The realized level of household indirect utility depends on the choice of a particular brand-size. If household  $h$  chooses brand-size  $j$  at  $t$



period which maximizes utility, then the level of utility is expressed as:

$$U_j^h(t) = V_j^h(t) + \varepsilon_j^h(t) = B'X_j^h(t) + \varepsilon_j^h(t)$$

where  $B$  is a coefficient vector and  $X_j^h(t)$  is the observable attribute vector (defined later).

We define an indicator variable  $\delta_{hj}(t)$  to denote choice  $j$  for household  $h$  at time  $t$ .

$\delta_{hj}^h(t) = 1$  if the  $h^{th}$  household chooses the brand  $j$  at  $t$  time

$\delta_{hj}^h(t) = 0$  otherwise

Assuming that the unobservable attributes are identically and independently distributed with a Type I extreme value(or Weibull) distribution,  $F(\varepsilon_m^h(t)) = \exp(-\exp(-\varepsilon_m^h(t)))$ , the probability of choosing a particular brand-size  $j$  for household  $h$  at  $t$  occasion is expressed as:

$$\begin{aligned} P_j^h(t) &= Pr(\delta_{hj}^h(t) = 1) = Pr(\varepsilon_m^h(t) < \varepsilon_j^h(t) + V_j^h(t) - V_m^h(t) \quad \forall m \neq j) \\ &= \int_{-\infty}^{\infty} \prod_{m \neq j} F(\varepsilon_j^h(t) + V_j^h(t) - V_m^h(t)) f(\varepsilon_j^h(t)) d\varepsilon_j^h(t) \end{aligned}$$

where  $F$  and  $f$  are cumulative and density distribution functions which are identically distributed with Type I extreme-value, respectively.

Following MacFadden's integral method (for details see MacFadden, 1974, pp. 106-113), household  $h$ 's choice probability of the  $j^{th}$  alternative has the simple form This expression is

$$P_j^h(t) = P(\delta_{hj}^h(t) = 1) = \frac{e^{V_j^h(t)}}{\sum_{m=0}^M e^{V_m^h(t)}} = \frac{e^{B'X_j^h(t)}}{\sum_{m=0}^M e^{B'X_m^h(t)}}.$$

known as a multinomial logit model. An important property in the multinomial logit model is the independence of irrelevant alternatives (IIA), which states that the odds of brand-size  $i$  being chosen over brand-size  $j$  is independent of the availability of attributes of alternatives other than  $i$  and  $j$ . This is an advantage since this property allows for the introduction of a new brand-size alternative without re-estimation of the model. This property is possible because the addition of a brand-size alternative does not change the relative odds with which the previous brand-size alternatives are chosen. The response probabilities for a brand-size choice in the expanded brand-size sets are obtained from the probability equation simply by adding terms in the denominator. However, this IIA property is a considerable restriction to place on household behavior.

Once the estimated parameter vector is obtained, the own and cross elasticities of choice probabilities for the brand-sizes can be calculated as:

$$\frac{\partial \ln P_j}{\partial \ln x_{jk}} = \beta_k x_{jk} (1 - P_j)$$

$$\frac{\partial \ln P_j}{\partial \ln x_{mk}} = - \beta_k x_{mk} P_m$$

$\tilde{P}_j$  is the overall estimated probability of choice for brand-size  $j$ . The interesting result is that the cross elasticity does not depend on  $j$ . This is due to the property of IIA and implies that all cross elasticities are equal.

### Nested Logit Model

The previous section described the multinomial logit model and pointed out the weakness of its IIA property. As an alternative, McFadden's nested multinomial logit model is used to estimate the household's choice probability of the brand-sizes.

In this section, we suggest a brand-size choice model that incorporates a decision-making process in which the choice alternatives are interdependent within choice clusters. In the first-step decision, the household chooses between purchasing or not purchasing spaghetti during a given week. The second-step decision is to decide which brand alternative to buy under the condition that the purchase alternative was chosen. The third-step decision is to decide which package size alternative to purchase under the condition that the brand was chosen.

Let household  $h$ 's indirect utility function in  $t$  time period be expressed by the linear form:

$$U_{ijk}^h(t) = V_{ijk}^h(t) + \varepsilon_{ijk}^h(t)$$

where  $i$  is the binary variable ( $i=1$  if purchased,  $i=0$  if not purchased),  $j$  is the brand  $j$  on a choice set  $C$  consisting of  $L$  brands,  $k$  is the size  $k$  on a choice set  $D$  consisting of  $M_j$  sizes in brand  $j$ ,  $V_{ijk}$  is a function of the measured characteristics of the household and the spaghetti products, and  $\varepsilon_{ijk}$  is a random variable that captures the effect of unmeasured variables from the researcher's point of view. We assume that  $\varepsilon_{ijk}$  are independently and identically distributed with a generalized extreme value.

Assume the measurable household indirect utility function has three component parts:

$$V_{ijk}^h(t) = B'X_{ijk}^h(t) + A'Y_{ij}^h(t) + \Gamma'Z_i^h(t),$$

where  $X_{ijk}$  is the vector of observed attributes including some demographic variables (defined later) that influence the third-step decision, i.e., which size to buy; where  $Y_{jk}$  is the vector of attributes including some demographic factors that influence the second-step decision, i.e., which brand to buy; where  $Z_k$  is the vector of attributes including some demographic factors that influence the first-step decision, i.e., whether to buy spaghetti or not; and B, A, and  $\Gamma$  are unknown parameter vectors.

The probability of choosing k size and j brand can be specified as follows:

$$P_{1jk}^h(t) = \frac{e^{\frac{B'X_{1jk}^h(t)}{(1-\lambda)(1-\rho)} + A'Y_{1j}^h(t) + \Gamma'Z_1^h(t) - \lambda IVI_{1j}^h(t) - \rho IVJ_1^h(t)}}{e^{\Gamma'Z_1^h(t) + (1-\rho) IVJ_1^h(t)} + e^{B'X_0^h(t) + \Gamma Z_0^h(t)}}.$$

$IVI_{ij}^h(t)$  is the inclusive value at the second-step decision defined as

$$IVI_{1j}^h(t) = \log \left( \sum_{m=1}^{M_i} e^{\frac{B'X_{1jm}^h(t)}{(1-\lambda)(1-\rho)}} \right).$$

$IVJ_1^h(t)$  is the inclusive value at the first-step decision defined as

$$IVJ_1^h(t) = \log \left( \sum_{l=1}^L e^{A'Y_{1l}^h(t) + (1-\lambda) IVI_{1l}^h(t)} \right).$$

The inclusive values IVI and IVJ provide a basis for identifying the behavioral relationship among brand alternatives and between size alternatives at each decision stage of the nested structure, as well as a test of the consistency of structure with utility maximization. The  $1-\rho_i$  and  $1-\lambda_i$  are scale factors at the first and second-step decisions respectively. These are the measure of the dependence of alternatives within the branch of decision steps. If the  $\rho_i$  and  $\lambda_i$  are zero, then

the generalized functional form is just a multinomial logit model.

The probability  $P_{ijk}$  can be expressed in terms of two conditional and marginal probabilities:

$$P_{1jk}^h(t) = P_{k|j,1}^h(t) P_{j|1}^h(t) P_1^h(t).$$

This means that the probability of choosing  $k$  size of  $j$  brand is the product of the conditional probability of choosing  $k$  size given the  $j^{th}$  brand choice and the purchase alternative was chosen, the conditional probability of choosing the  $j^{th}$  brand given the purchase alternative was chosen, and the marginal probability of choosing the purchase alternative.

Two conditional and marginal probabilities can be derived as:

$$(1) P_{k|j,1}^h(t) = \frac{e^{\frac{V_{1jk}^h(t)}{(1-\lambda)(1-\rho)}}}{\sum_{m=1}^{M_j} e^{\frac{V_{1jm}^h(t)}{(1-\lambda)(1-\rho)}}} = \frac{e^{\frac{B'X_{1jk}^h(t)}{(1-\lambda)(1-\rho)}}}{\sum_{m=1}^{M_j} e^{\frac{B'X_{1jm}^h(t)}{(1-\lambda)(1-\rho)}}}.$$

$$(2) P_{j|1}^h(t) = \frac{e^{A'Y_{1j}^h(t) + (1-\lambda)IVI_{1j}^h(t)}}{\sum_{l=1}^L e^{A'Y_{1l}^h(t) + (1-\lambda)IVI_{1l}^h(t)}}.$$

$$(3) P_1^h(t) = \frac{e^{\Gamma'Z_1^h(t) + (1-\rho)IVJ_1^h(t)}}{e^{\Gamma'Z_1^h(t) + (1-\rho)IVJ_1^h(t)} + e^{B'X_0^h(t) + \Gamma'Z_0^h(t)}}.$$

Equation (1) represents the household's third-step decision probability (i.e., which size to buy

after brand  $j$  and the purchase alternative are chosen). Equation (2) is associated with the second-step decision behavior of the household (i.e., which brand to choose after deciding to buy).

Equation (3) indicates the first-step decision behavior of the household (i.e., whether to buy spaghetti or not).

The elasticities of choice probabilities can be derived from a method similar to that for the MNL model. Since the joint probability  $P_{ijk}$  is indirectly estimated from two conditional probabilities,  $P_{k|j,1}$  and  $P_{j|1}$ , and the marginal probability  $P_1$ , the own and cross elasticities of attributes are calculated as:

$$\begin{aligned} \frac{\partial P_{1lm}}{\partial X_{1jk,s}} \cdot \frac{X_{1jk,s}}{P_{1lm}} = & \beta_s \cdot X_{1jk,s} [\delta_{km} - (1-\lambda)(1-\rho) \cdot P_{k|j,1} \cdot P_{j|1} \cdot P_1 \\ & + (1-\lambda)(-\rho) P_{k|j,1} \cdot P_{j|1} + (-\lambda) \cdot \delta_{jl} \cdot P_{k|j,1} \end{aligned}$$

where  $\delta_{km}$  is the Kronecker symbol ( $\delta_{km}=1$  if  $k=m$ ,  $\delta_{km}=0$  otherwise) and  $X_{1jk,s}$  is the mean of the  $j^{\text{th}}$  brand and  $k^{\text{th}}$  size of the  $s^{\text{th}}$  attribute. These elasticities measure the percentage change of the probability of choosing the  $l^{\text{th}}$  brand and  $m^{\text{th}}$  size alternative in response to a change in the  $s^{\text{th}}$  attribute of the  $j^{\text{th}}$  brand and  $k^{\text{th}}$  size alternative by one percent.

## DATA AND VARIABLES

The A.C. Nielsen household panel scanner data on the purchases of regular spaghetti in the nationwide market were used. We selected the first quarter of 1994 as the sample period. Twenty-two grocery store chains were selected based on data availability for market information on price and promotion for the brand-size alternatives. Seven choice alternatives were selected: six brand-size alternatives and a no-purchase alternative. For brand-size choices, 16 oz. and 32 oz. of Creamette were first selected because it holds the largest market share nationally and, excepting Creamette, the 16 oz. and 32 oz. packages of the brands with the first and second largest market share in each grocery store were used.

The price and sales activity for these brand-size alternatives on a store-level was obtained from the household purchase records on a weekly basis. If it was the case that no panelist purchased one of the alternatives during a particular week, we used the previous week's price as the current week's price. For market environment, such as advertising and promotion of that brand, we assumed there was no special promotional activity going on at that time.

From the 22 supermarket chains, in all, 1,744 households made 2,877 purchases of the six alternatives during the first quarter of 1994. We randomly chose 1,000 households which had purchased the six brand-size alternatives at least once during the 1994 first quarter (13 weeks). In the sample selection, households which had purchased different alternatives at the same time were eliminated because this behavior violates the assumption that households choose only one alternative which would maximize their utility.

With the 1,000 household sample, there was a total of 91,000 observations (1,000

households x 13 weeks x 7 alternatives). In the sample (1,000 households x 13 weeks), 1,526 (12 percent) purchased one of the brand-size alternatives, while 11,474 (88 percent) had chosen the no-purchase alternative during a given week.

### Explanatory Variables

The summary statistics for attribute variables and the definition of demographic variables and their descriptive statistics are presented in Table 1 and Table 2 respectively.

Table 1. Descriptive Statistics for Attribute Variables<sup>a</sup>

Variable	Choice 1	Choice 2	Choice 3	Choice 4	Choice 5	Choice 6	Choice 7
PRICE <sup>b</sup>	66.89 (7.65)	129.80 (13.83)	70.69 (11.29)	139.60 (6.42)	93.43 (6.99)	180.07 (9.38)	NA <sup>e</sup>
PRICECUT <sup>b</sup>	0.449 (2.80)	0.533 (3.61)	1.068 (6.54)	0.047 (1.22)	0.853 (5.69)	0.673 (7.64)	NA
LOYALTY <sup>c</sup>	0.128 (0.33)	0.079 (0.27)	0.176 (0.38)	0.058 (0.23)	0.097 (0.29)	0.043 (0.20)	NA
BLOYALTY <sup>c</sup>	0.207 (0.40)	0.207 (0.40)	0.233 (0.42)	0.233 (0.42)	0.142 (0.34)	0.141 (0.34)	NA
ADVER <sup>c</sup>	0.038 (0.19)	0.021 (0.14)	0.016 (0.12)	0.003 (0.05)	0.020 (0.14)	0.0006 (0.02)	NA
INVENT <sup>d</sup>	NA	NA	NA	NA	NA	NA	8.03 (10.25)

<sup>a</sup>Means and standard deviations in parentheses.

<sup>b</sup> Cents per package.

<sup>c</sup> Binary variables described in the text.

<sup>d</sup> Measured in ounces.

<sup>e</sup> Not available. However, zeros were used in estimation.



Table 2. Definitions of Demographic Variables and Descriptive Statistics

Variable	Definition	Mean	S.D
INCOME	Family income category: 1 if under \$9,999 2 if between \$10,000 and \$19,999 3 if between \$20,000 and \$29,999 4 if between \$30,000 and \$39,999 5 if between \$40,000 and \$49,999 6 if between \$50,000 and \$59,999 7 if between \$60,000 and \$69,999 8 if over \$70,000	4.262	1.920
HHSIZE	Number of members in household h	2.835	1.387
WHITE	1 if household h is non-Hispanic white; 0 otherwise	0.884	0.320
HISP	1 if household h is Hispanic	0.039	0.193
MARRIED	1 if married	0.720	0.449
AGE1	1 if the age of the female head of the household (or the age of male head of the household if there is no female head) is under 34	0.171	0.376
AGE2	1 if the age of the female head of the household (or the age of male head of the household if there is no female head) is between 35 and 65	0.589	0.492
CHILD	1 if household h has children under 18; 0 otherwise	0.342	0.475
FEMPLOY	1 if the female head of household h has a full time job	0.329	0.469
COUNTY1	1 if household h is in a county belonging to one of the 25 largest standard consolidated statistical areas (SCAS) or standard metropolitan statistical area (SMSA)	0.433	0.495
COUNTY2	1 if household h is in a county that is a SCAS or SMSA with over 150,000 population and does not belong to COUNTY1	0.315	0.464
EAST	1 if household h is located in the East	0.283	0.450
SOUTH	1 if household h is located in the South	0.288	0.452
CENTRAL	1 if household h is located in the Central region	0.368	0.482
EMH	1 if the male head of household h has at least 1 year of college	0.304	0.459
EFH	1 if the female head of household h has at least 1 year of college	0.247	0.431
OCCUP	1 if the head of household h has a professional or white collar job	0.415	0.492

In Table 1, Choice 1 is the 16 oz. size of the lower priced brand with the first or second largest share in the supermarket, excluding Creamette; Choice 2 the 32 oz. size; Choice 3 is the higher priced brand with the first or second largest share excluding Creamette; Choice 4 the 32 oz. size; Choice 5 is the 16 oz. size of Creamette; Choice 6 the 32 oz. size; and Choice 7 is the no-purchase alternative.

### Multinomial Logit Model

The variables assumed to affect the household's choice of brand-size alternatives are constant terms, prices (PRICE), price-cuts (PRICECUT), advertising features (ADVER), loyalty (LOYALTY), household inventory levels (INVENT), household income (INCOME), and household size (HHSIZE).

Brand-size constant terms are created by using a set of dummy variables to represent any uniqueness of an alternative that is not captured by other attribute variables as well as demographic factors. PRICE and PRICECUT are per package values expressed in cents. PRICE is the actual price paid by a household. PRICECUT is the amount of any price cut and is zero when there is no discount on a brand-size alternative. ADVER is a dummy variable equal to one if there is an advertising feature that was not accompanied by a price reduction and zero otherwise. INVENT is recursively calculated by the following method:  $INVENT = INVENT_{t-1} + Q_{t-1} + DR$  where  $INVENT_{t-1}$  is last week's inventory level,  $Q_{t-1}$  is the total ounces of product bought the previous week and is zero if the product was not purchased, and DR is the weekly discount rate which is computed as the total number of ounces purchased by household h in the first quarter 1994 divided by 13 weeks in the sample period.  $INVENT_{t-1}$  and  $Q_{t-1}$  were set at zero at the beginning of the sample period. Since the household's inventory level is common across alternatives, we treat this variable as a unique attribute affecting Choice 7 (the no-purchase

alternative), and set it equal to zero for the other six alternatives, as can be seen in Table 1.

We add a LOYALTY variable in the spirit of Guadagni and Little (1983). LOYALTY is set equal to one if the brand-size was purchased on the last purchase occasion at time  $t-i$  and zero otherwise. LOYALTY and INVENT are considered state variables in this model. LOYALTY can be interpreted as a psychological stock in the sense that it represents the household's brand-size purchasing habit or "inertia", while INVENT represents the physical stock which is the household's current stock-adjustment factor (Houthakker and Taylor, 1970).

### Nested Logit Model

The first-step equation represents the probability of purchasing or not at a particular time among the consumers who have purchased one of the spaghetti alternatives during the sample period. The variables assumed to affect the household's choice in the first-step equation are a constant term, INVENT, the inclusive value (IVJ) calculated from the second stage estimation equation, and demographic factors. The household panel data include demographic information only for households which purchase a product.

In the second-step decision equation, constant terms, brand loyalty (BLOYALTY), inclusive value (IVI) calculated from the first-stage estimation equation, and demographic factors are the explanatory variables. The constant terms of each brand are created using a set of dummy variables. The constant will capture the uniqueness of a brand that is not revealed by other variables. BLOYALTY is measured in the same method as LOYALTY. Demographic factors that might affect tastes and preferences as determinants of the probability of choosing a certain brand are used.

The explanatory variables used in the third-step equation are the same as the variables specified in the (non-nested) multinomial logit model including price, price cut, promotion, brand

loyalty, and household income and size. Constant terms of sizes rather than those of brand-size are used in this equation.

## EMPIRICAL RESULTS

In this section we present the results and discuss their implications for the two different models. The first is the nested logit model which sequentially reduces the choice set, and the second is the multinomial logit model using the entire choice set.

### Nested Logit Model

In the estimation procedure, equations were estimated by a three-stage sequential method which provides consistent parameter estimates (Maddala, 1983, p. 70). In the first-stage estimation, the household's third-step decision equation (which size to buy) was estimated and the inclusive value IVI was calculated from the estimated equation. In the second-stage estimation, the household's second-step decision equation (which brand to buy) was estimated by including the inclusive value IVI which became a regressor in the second stage and the inclusive values IVJ was computed from the estimated equation. In the third-stage estimation, the household's first-step decision equation (to buy spaghetti or not in a given week) was estimated by including the inclusive value IVJ which became a regressor in the third-stage estimation.<sup>2</sup>

The results for sequential maximum likelihood estimation of the nested multinomial logit equations are given in Table 3. The estimated chi-squared values indicate that the equations are highly significant. The equations correctly predict 58.8, 33.1, and 65.7 percent of the observations, respectively.

Table 3. The Estimation Results for the Decision Equations

First-Step Equation Prob. of purchase/not		Second-Step Equation Prob. of brand choice j		Third-Step Equation Prob of size choice k	
Variable Coefficients <sup>a</sup>		Variable Coefficients		Variable Coefficients	
INTERCEPT	-2.0718* (-12.08)	INTERCEPT1 <sup>b</sup>	-0.22531 (-0.073)	INTERCEPT1 <sup>c</sup>	-0.89103* (-2.022)
IVJ	0.85509* (7.853)	INTERCEPT2	-0.24948 (-0.120)	INTERCEPT3	-0.52594 (-0.966)
INVENT	-0.03202* (-7.006)	IVI	0.78047* (5.767)	INTERCEPT5	-1.4480* (-2.435)
WHITE	0.18768 (1.701)	EMH1	-0.14285 (-0.007)	PRICE	-0.02860* (-5.846)
HISP	0.34340* (2.075)	EMH2	-0.06764 (-0.005)	PRICECUT	0.03019* (2.212)
AGE1	0.06657 (0.678)	EFH1	-0.09262 (-0.012)	ADVER	1.1489* (3.006)
AGE2	0.00682 (0.093)	EFH2	0.11055 (0.022)	LOYALTY	1.9371* (10.05)
CHILD	0.21754* (3.356)	BLOYALTY	0.01227 (0.186)	INCOME1	-0.09610 (-1.625)
MARRIED	0.00881 (0.135)	OCCUP1	-0.02505 (-0.006)	INCOME3	0.01126 (0.187)
FFEMPLY	-0.12027* (-1.961)	OCCUP2	-0.09594 (-0.035)	INCOME5	-0.02344 (-0.308)
COUNTY1	0.01416 (0.194)			HHSIZE1	-0.04884 (-0.608)
COUNTY2	0.11912 (1.628)			HHSIZE3	-0.17146 (-1.944)
CENTRAL	-0.36761* (-3.088)			HHSIZE5	0.00672 (0.074)
SOUTH	-0.04185 (-0.348)				
EAST	-0.09526 (-0.809)				
Chi-Squared	10614.13		1148.54		1436.71

<sup>a</sup>Asymptotic t-statistics are reported in parentheses.

\*Indicates statistical significance at at least the 5% level, with a t-statistic  $\geq 1.96$ .

<sup>b</sup>The numbers (1, 2) relate to alternative brands.

<sup>c</sup>The numbers (1, 3, 5) relate to package size for each brand.

Since the data in our model were households which purchased the brand-size alternatives at least once during the sample period, the first-step equation represents the probability of purchase among the consumers who have purchased these alternatives. The results show that INVENT is a significant factor in decision making. The more inventory left, the less likely a household is to buy spaghetti in a particular week.

Also in the first equation in Table 3, the CHILD variable has a significant positive effect on the buying alternative. The FEMPLOY effect is significantly negative in this equation, implying that full-time working women tend to purchase spaghetti products less frequently. Two racial groups have a positive influence on the purchase alternative, although for whites the coefficient is significant at only the 10 percent level. Since the intercept term in the estimated equation partially represents the effects of the racial groups not specified, positively significant coefficients imply that whites and Hispanics purchase spaghetti more frequently than other racial groups such as African-Americans and Asians. Residents of the Central region buy spaghetti less often than those in the West.

The sequential second-step decision equation was estimated using brand alternatives as a dependent variable. Since household demographic variables were a common attribute to the choice alternatives, we created a set of dummy variables for the choice alternatives and multiplied each of them by the common demographic factor so that parameters of the demographic factor could be estimated (for details, see MacFadden, 1984). Demographic variables for the third brand choice (Creamette) were deleted in order to avoid the dummy variable trap (perfect colinearity). The results, however, indicate that none of the demographic variable (income and household size) help to explain household brand choice behavior.

The estimated coefficients of IVJ and IVI represent the dissimilarity factors at each stage.

If the dissimilarity factor is equal to one, it means that choice alternatives are completely dependent within the choice set at the decision stage. If it equals zero, choice alternatives are independent within the choice sets. The nested structure then collapses to the multinomial logit model that has the property of independence of irrelevant alternatives. The results of dissimilarity tests suggest that households may not follow the set of recursive decision steps implied by the nested model and that the (non-nested) multinomial logit model is appropriate. See Appendix A.

The third-step equation in Table 3 was estimated using two size choices, 16 oz. and 32 oz., for each brand as dependent variables. In estimation, demographic variables for 32 oz. were taken as the reference points. The estimated equation shows that all marketing variables are very important for a household's brand-size choice decisions. The coefficients of all market environment variables such as price, price-cut, advertising (ADVER), and loyalty have the algebraic signs that would be expected and are statistically significant at the 5 percent level or better.

Loyalty, which represents household habit formation, is a measure of the brand's repeat purchase incidence. Households which purchase a certain brand-size alternative are more likely to repurchase that alternative next time under the same market circumstances. Empirical studies (Massy, 1966 and Bass et.al, 1984) indicated that consumer brand choice was consistent with a "no learning" process, especially in cases of inexpensive nondurable consumer goods. In this case also, households are more likely to be using a buying process which includes making routine choices or by applying very simple choice habits.

Since different brand-sizes of spaghetti on the same supermarket shelf are very close substitutes and have similar attributes, the effect of a price difference between alternatives is large and is a major factor in either brand-size switching or repeat purchasing. Price-cuts, which are a common sales promotion technique, have a significant positive effect. It is almost the same as the price effect but in the opposite direction since the existence of a price-cut is specified as one and

zero otherwise. Households seem to treat price-cuts as another way of obtaining price reductions. Advertising is also an important factor in household response, increasing the probability of a brand-size purchase.

The utility structure of a household was assumed to be a linear function of choice alternatives in our model, which implies that income should be positively correlated with the probability of purchasing a product. That was not the case in this study; however, none of income variables are statistically significant. There are two main reasons for this. First, expenditures on spaghetti are, in general, a very small portion of household income so that increasing household income does not significantly change the probability of the choice of spaghetti. Second, the brand-size products used in our study are the lower quality spaghetti products (Creamette and the two other best selling brands in the store which are the lower price ones). Therefore, the income variables probably have less influence on brand-size alternatives than if some more expensive, premium brands were included.

With the hypothesis that larger households are less likely to choose smaller package sizes the expected signs of the household size variables were all negative. Two of the three variables have the correct sign, but only household size for Choice 3 is significant at the 10 percent level indicating that as household size increases the probability of buying the 16 oz. size of Brand 2 decreases.

Price elasticities of probability are calculated from the formula explained previously. These are analogous to normal price elasticities but relate to the choice probability related to price changes, as opposed to quantity consumed. Specifically, the own-price elasticities of choice probability indicate the percent change in the probability of a choice with respect to a one percent change in that price. Previous research (Blanciforti and Green, 1983; Falconi and Senauer, 1991; Huang, 1985) shows that the estimated own-price elasticities for food products are usually



inelastic, with most between 0 and -0.5. However, the estimated own-price elasticities of probability (Table 4) indicate that all brand-size choices are highly elastic. The major reason is that these alternatives are sufficiently similar to be very close or nearly perfect substitutes for one another so that households are very sensitive to price differences.

Table 4. The Estimated Own and Cross Elasticities of Choice Probabilities with Respect to Prices based on the Nested Logit Model

	Brand 1		Brand 2		Brand 3	
	Choice 1	Choice 2	Choice 3	Choice 4	Choice 5	Choice 6
PRICE 1	-1.56504	<b>.34801</b>	.07388	.07388	.07388	.07388
PRICE 2	<b>.35929</b>	-3.35300	.00747	.00747	.00747	.00747
PRICE 3	.15738	.15738	-1.48933	<b>.53240</b>	.15738	.15738
PRICE 4	.05702	.05702	<b>.19293</b>	-3.79962	.05702	.05702
PRICE 5	.08436	.08436	.08436	.08436	-2.15344	<b>.51863</b>
PRICE 6	.05703	.05703	.05703	.05703	<b>.35063</b>	-4.79939

Choice 6 has the largest own-price responsiveness with -4.79939, while Choice 1 has the smallest with -1.56504. The 32 oz. alternatives (Choices 2, 4 and 6) have larger price elasticities than the 16 oz. sizes (Choices 1, 3 and 5).

All cross elasticities in Table 4 show larger price responsiveness within the same brand between package sizes (bold numbers) than between brands. Note that the cross-elasticities beyond the same brand are the same across the brand-size alternatives due to the fact that the choice alternatives outside the brand are independent of each other as discussed previously. Brand 1 has almost the same cross effect between the two size alternatives. For example, if the price of Choice 1 (the 16 oz. size) increases one percent then the probability of purchasing Choice 2 (the 32 oz. size) increases by .34801 and if the price of Choice 2 increases one percent, the

probability of purchasing Choice 1 increases by .35929.

### Multinomial Logit Model

In the last section, the two similarity tests for the inclusive values (IVI and IVJ) showed that the tightness of choice cluster at each decision step was not strong and that the NMNL (nested) model was weakly rejected against the null hypothesis of the MNL (non-nested) specification. In this section, we present the results of the multinomial logit model which estimates the probability of choosing any one of the seven alternatives (the six brand-size choices and the no purchase option). In this model, only two demographic variables INCOME and HH SIZE are included in the estimated equation, so the two models (NMNL and MNL) are not directly comparable. The demographic variables had to be limited because each must be separately specified for the seven choices (the six brand-size alternatives and the no-purchase choice). The MNL model was estimated by a maximum likelihood method using the Newton-Raphson approach.

The Chi-squared value in Table 5 indicates that the equation is highly significant. The model correctly predicts 75.9 percent of the observations. All marketing variables (price, price-cut, and advertising), loyalty, and inventory (INVENT) have the expected signs and are significant at the one percent level. Brand-size loyalty and household inventory are the most important factors in determining the choice alternatives. The estimated coefficients with the largest magnitude are for the loyalty and advertisement variables which were dummy variables and shift the intercepts. There is a strong negative inventory effect on the decision to buy spaghetti. If household inventory stocks are larger during the current week, the choice of one of the brand-size alternatives is less likely for this period.

The brand-size constants form a special factor. They can represent unique alternative

attributes or specification errors. If the other explanatory elements perfectly explain households choice behavior, these constants should be zero. Choice 2 has been taken as the reference point and is deleted in order to avoid the dummy variable trap. Compared to other dummy variables such as LOYALTY and ADVER, most of the estimated constant terms are small and statistically zero (not significant), except Choice 7. The explanatory variables for brand-size alternatives appear to explain the household response to brand-size alternatives, while the INVENT and demographic variables are not enough to completely explain the household's decision not to purchase.

Estimated income coefficients for brand-size alternatives have mixed signs and are generally not significant: four out of the six coefficients are positive while Choice 1 and Choice 4 have negative signs. Only Choice 3 has a statistically significant and positive sign which means that the probability of choosing Brand 2 in the 16 oz. size increases as income increases. The expected sign of the coefficients for household size were negative for 16 oz. sizes and positive for 32 oz. sizes. As the number of household members increase, households were expected to shift their choices from smaller to large packages. The estimated results show that all household size coefficients are negative. Choice 1 and Choice 3 (16 oz. sizes) have the correct signs and are statistically significant while the other choices are not significant. This implies that the customers who choose the 16 oz. size of the leading brands are likely to have a smaller number of household members.

Table 5. The Estimation Results of the Multinomial Logit Equation

Variables	Coefficients <sup>a</sup>
INTERCEPT1	-0.28327 (-0.816)
INTERCEPT3	-0.11142 (-0.332)
INTERCEPT4	0.14661 (0.822)
INTERCEPT5	-0.38424 (-1.215)
INTERCEPT6	0.34769 (0.860)
INTERCEPT7	2.8635* (5.797)
PRICE	-0.02014* (-6.297)
PRICECUT	0.01262* (3.165)
ADVER	1.0627* (8.280)
LOYALTY	1.8070* (13.410)
INVENT	-0.03331* (-12.646)
INCOME1	-0.04541 (-0.883)
INCOME3	0.09621* (2.016)
INCOME4	-0.00273 (-0.083)
INCOME5	0.01302 (0.244)
INCOME6	0.00793 (0.111)
INCOME7	0.03623 (0.849)
HHSIZE1	-0.10274 (-1.617)
HHSIZE3	-0.21016* (-3.457)
HHSIZE4	-0.01757 (-0.397)
HHSIZE5	-0.05486 (-0.827)
HHSIZE6	-0.05912 (-0.660)
HHSIZE7	-0.14173* (-2.695)
Chi-squared	27109.02

<sup>a</sup>Asymptotic t-statistics are reported in parentheses.

\*Indicates statistical significance at at least the 5% level, with a t-statistic  $\geq 1.96$ .

The estimated own-price elasticities of choice probability (Table 6) show that all brand-size choices are highly elastic. Choice 6 (Creamette 32 oz. size) has the highest own-price response with -3.60770, while Choice 1 has the lowest price response with -1.28682. The higher priced brand-size has a higher price elasticity than the lower priced brand-size alternatives. Also the 32 oz. sizes have higher price elasticities than the 16 oz. sizes.

The cross elasticities indicate the brand-size switching resulting from price changes. Note that the cross elasticities in each row in Table 6 are the same across brand-size alternatives and are asymmetric due to the independence of irrelevant alternatives (IIA) property in a multinomial logit model. For example, a one percent increase in the price of brand-size Choice 1 will increase the probability of choosing Choice 2 as well as other alternatives by .06060 percent, while a one percent increase in price of Choice 2 will increase the probability of choosing Choice 1 as well as other alternatives by .04408 percent. This IIA property is a considerable restriction to place on household choice behavior.

Table 6. The Estimated Own and Cross Elasticities of Choice Probabilities with Respect to Prices based on the Multinomial Logit Model

	Brand 1		Brand 2		Brand 3	
	Choice 1	Choice 2	Choice 3	Choice 4	Choice 5	Choice 6
PRICE 1	-1.28682	.06060	.06060	.06060	.06060	.06060
PRICE 2	.04408	-2.57061	.04408	.04408	.04408	.04408
PRICE 3	.07610	.07610	-1.34787	.07610	.07610	.07610
PRICE 4	.03030	.03030	.03030	-2.78180	.03030	.03030
PRICE 5	.04886	.04886	.04886	.04886	-1.83318	.04886
PRICE 6	.01962	.01962	.01962	.01962	.01962	-3.60770

## CONCLUSIONS

The main objectives of this study were to analyze the impact of market environment and household demographic factors on household choice behavior for a specific brand-size product. The A.C. Nielsen panel scanner data on the purchases of regular spaghetti in the nationwide market were used for the estimation of the brand-size choice model parameters.

In the nested logit model, estimation of the first-step decision model indicated that a larger household inventory of spaghetti decreased the probability of a purchase. While households with a female head with a full-time job were less frequent purchasers of spaghetti products, households with children were more frequent purchasers. Two racial groups, whites and Hispanics, had a positive and significant influence on overall spaghetti purchases. Estimation of the second-step equation showed that the demographic factors failed to explain household choice of brand. In the third-step decision model, all market variables such as price, price-cut, advertising, plus loyalty were very important for explaining the brand and size of spaghetti chosen by households.

Estimation of the (non-nested) multinomial model also showed that market environment variables such as price and price promotion are important factors in determining which brand-size to choose. The current stock of inventory had a strong negative effect on the probability of buying spaghetti in a given week. Income was generally not a significant variable. This result is reasonable given that this study used lower price and standard quality spaghetti products. Creamette and the two other best selling brands in a store were utilized, which excluded smaller market share, premium brands that higher income households might be more likely to buy.

The most interesting results found in this study relate to the price elasticities of choice probabilities. The price elasticity of choice probability shows the percent change in choice probability with respect to a one percent change in price. From the two models, the estimated own-price elasticities of choice probabilities indicated that all brand-size alternatives were highly

elastic. Different brands and sizes of spaghetti products on the same grocery store shelf are much closer substitutes than more aggregate food categories, such as beef, pork, and poultry, used in traditional food demand analysis. The results in this study are consistent with microeconomic theory which suggests that the more substitutes a product has and the closer they are, the more elastic will be the response to price changes.

The models used in this study could be extended in a number of directions. First, assuming that the utility structure was linear and homothetic implies that all brand-size alternatives are normal goods. Further studies could ease this assumption by making the utility structure a nonhomothetic function which is linear but has a changing slope for the indifference curves as the level of household utility increases. Choice would be allowed to switch from a lower quality to a higher quality brand-size rather than purchasing more of the same brand-size. Second, the factors determining households which never purchase spaghetti were not analyzed. This study could be extended by considering households which never purchase the product.

The availability of scanner data linked to households and supermarkets poses new issues and problems for demand analysis. However, such data open rich new possibilities to improve our understanding of consumer behavior.

## REFERENCES

- Amemiya, T. "On a Two-Step Estimation of a Multivariate Logit Model." *Journal of Econometrics* 8(1978):13-21.
- . *Advanced Econometrics*. Cambridge. Harvard University Press. 1985.
- Bass, F. M., M. M. Givon, M. U. Kalwani, D. Reibstein, and G. P. Wright. "An Investigation into the Order of the Brand Choice Process." *Marketing Science* 3(1984): 267-287.
- Blanciforti, L., and R. Green. "The Almost Ideal Demand System: Comparison and Applications to Food Groups." *Agri. Econ. Res.* 35(1983): 1-10.
- Bucklin, R. E., and J. M. Lattin. "A Two-State Model of Purchase Incidence and Brand Choice." *Marketing Science* 10(1991): 24-39.
- Bunch, K. and B. Wendland. "Grain Products Regain Popularity." *National Food Review*. Washington, D.C.: U.S. Dept. of Agriculture, ERS, NFR-34, Summer 1986.
- Capps, O., Jr. "Utilizing Scanner Data to Estimate Retail Demand Functions for Meat Products." *Amer.J.Agr.Econ.* (1989): 750-760.
- Capps, O., Jr., and J. A. Lambregts. "Assessing Effects of Prices and Advertising on Purchases of Finfish and Shellfish in a Local Market in Texas." *S. J. of Agri. Econ.* (1991): 181-194.
- Capps, O., Jr., and R. M. Nayga, Jr. "Demand for Fresh Beef Products in Supermarkets: A Trial with Scanner Data." *Agribusiness* 7(1991):241-251.
- Daly, A., and S. Zachary. "Improved Multiple Choice Models." *Identifying and Measuring the Determinants of Mode Choice*. D. Hensher and Q. Dalvi, eds., pp. 335-57. London: Teakfield, 1979.
- Debreu, G. "Review of R.D. Luce Individual Choice Behavior." *Amer. Econ. Rev.* 50(1960): 186-8.
- Fader, P. S., and L. McAlister. "An Elimination by Aspects Model of Consumer Response to Promotion Calibrated on UPC Scanner Data." *J. of Marketing Research* 28(1990): 322-32.
- Greene, W. H. *Econometric Analysis*. New York. MacMillan Publishing Company. 1990.
- Guadagni, P. M., and J. D. Little. "A Logit Model of Brand Choice Calibrated on Scanner Data." *Marketing Science* 2(1983): 203-238.
- Hanemann, M. W. "Discrete/Continuous Model of Consumer Demand." *Econometrica* 52(1984): 541-561.



- Houthakker, H. S., and L. D. Taylor. *Consumer Demand in the United States: Analyses and Projections*. Cambridge. Harvard University Press. 1970.
- Huang, K. "U.S. Demand for Food: A Complete System of Price and Income Effects." USDA, ERS, Technical Bulletin No. 1714, Washington, D.C., 1985.
- Maddala, G. S. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge. Cambridge University Press. 1983.
- Massy, W. F. "Order and Homogeneity of Family Specific Brand-Switching Processes." *J. of Marketing Research* 3(1966): 48-54.
- McFadden, D. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, ed. by P. Zarembka. New York. Academic Press. 1973.
- \_\_\_\_\_. "Quantal Choice Analysis: A Survey." *Annals of Economic and Social Measurement* 5(1976): 363-390.
- \_\_\_\_\_. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data with Econometric Applications*, ed. by Manski, C.F., and D. McFadden. Cambridge. MIT Press. 1981.
- \_\_\_\_\_. "Econometric Analysis of Qualitative Response Models." In *Handbook of Econometrics*, ed. by Griliches, Z., and M. D. Intriligator. New York. North-Holland. 1984.
- Putman, J. J. and J. E. Allshouse. *Food Consumption, Prices, and Expenditures, 1970-1993*. Washington, D.C.: U.S. Dept. of Agriculture, ERS, Statistical Bulletin 915, December 1994.
- Russell, G., and W. A. Kamakura. "Understanding Brand Competition Using Micro and Macro Scanner Data." *J. of Marketing Res.* 31(1994): 289-303.
- Senauer, B., E. Asp, and J. Kinsey. *Food Trends and the Changing Consumer*. St. Paul. Eagan Press. 1991.
- Wagner, U., and A. Taubes. "A Multivariate Polya Model of Brand Choice and Purchase Incidence." *Marketing Science* 5(1986): 219-44.

## Endnotes

1. A linear utility structure, which is one disadvantage of the random utility model, does not allow both inferior and superior alternatives. All choice alternatives are normal goods. As income increases, the consumer demands more of the same alternative rather than substituting other higher quality alternatives.

2. Each equation was estimated by a maximum likelihood method using the Newton-Raphson approach. The ML method produces consistent estimators but is inefficient, except for the first-stage estimation, because in the second and third-stage estimation, variance-covariance matrices are affected by the use of estimated coefficients in the calculation of inclusive values (Amemiya,1978). In order to get efficient estimators, we followed McFadden's (1981) method which provides a recursive formula for obtaining the asymptotically efficient variance-covariance matrix at each stage.

## APPENDIX A

The test statistic for independence within the choice alternative is:

$$W = (IV-1) [var(IV)]^{-1} (IV-1)' \sim \chi^2$$

This statistic is a Wald statistic for the null hypothesis that the conditional logit is incorrect (i.e.,  $IV = 0$ ), which is Chi-square with one degree of freedom under null. The statistical values at the first and the second stage are 1.77 and 2.62 respectively. We reject the null hypothesis, at most, only at a 20 percent significance level at each stage. In the two dissimilarity tests, we conclude that the tightness of the choice cluster at each decision step is less strong. It implies that the household may not necessarily follow a series of decision steps, i.e., make choices in a certain order.

The test of the nested logit model with the random utility maximization (RAM) hypothesis can be established from two estimated inclusive values. These values should lie in the unit interval to be globally consistent with random utility maximization (for details, see Daly and Zachary, 1979 or MacFadden, 1981). Also, the magnitude of the estimated inclusive value in the upper level should be larger than that in the lower level. Based on the two estimated inclusive values, the household choice model is consistent with RAM.