8-2002

Determinants of Consumer Behavior in an e-Commerce Environment

Xiang Xue

DETERMINANTS OF CONSUMER BEHAVIOR IN AN E-COMMERCE ENVIRONMENT

By

Xiang Xue

B.A. Nankai University, 1999

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

(in Resource Economics and Policy)

The Graduate School

The University of Maine

August, 2002

Advisory Committee:

Gregory K. White, Associate Professor of Resource Economics and Policy, Advisor

Hsiang-Tai Cheng, Associate Professor of Resource Economics and Policy

Timothy J. Dalton, Assistant Professor of Resource Economics and Policy

LIBRARY RIGHTS STATEMENT

DETERMINANTS OF CONSUMER BEHAVIOR IN AN E-COMMERCE ENVIRONMENT

By Xiang Xue

Thesis Advisor: Dr. Gregory K. White

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Master of Science
(in Resource Economics and Policy)
August, 2002

Online specialty food and beverage marketing has developed rapidly in the last decade. With the obvious increase in the sales income, researchers showed more and more interests in this promising market. Previous studies in this field primarily focused on the demographic profiling of the online specialty food consumers as well as initial analysis in the factors that motivate their purchase behavior. However, it is far from fully explaining consumer's online shopping activities.

In order to solve these problems, new methods should be attempted to determine the factors influencing consumer's online buying behavior. A cluster analysis was developed to identify what kind of consumers have a higher preference of shopping online. And this analysis is based on consumer's lifestyle characteristics, which are assumed to be the determinants of their online shopping behavior. By identifying the consumer groups with a relatively high propensity of purchasing on the Internet, this

research will provide improved information for companies marketing food and beverage products online and strengthen their abilities to develop effective marketing strategies.

The data for this study were obtained from an online survey conducted from September 27th to October 25th, 2001. First a factor analysis was applied to the forty-two lifestyle variables and ten factors were extracted to represent them. Then the factor scores for each observation, an output data set of factor analysis, were submitted to a hierarchical cluster analysis using Ward's minimum-variance clustering method. As a result, lifestyle segmentations of the online specialty food and beverage market were obtained. Each segment was given a name based on the factor that has the highest loading on the particular group so as to emphasize its characteristics. Further analysis has also been done in order to compare the other related characteristics among the groups.

ACKNOWLEDGEMENTS

Many people deserve acknowledgements for their contribution to my education and the writing of this thesis. Dr. Gregory K. White, my major thesis advisor, deserves special recognition for the many hours he spent reviewing and constructively criticizing the work presented here. The other members of my thesis committee, Dr. Hsiang-tai Cheng and Dr. Timothy J. Dalton also contributed substantially to this endeavor and made it possible. To these three, I offer my sincere thanks for their invaluable assistance.

I would also like to thank the entire faculty, staff, and particularly my fellow graduate students of the Department of Resource Economics and Policy, University of Maine for making the time I have spent here truly enjoyable and memorable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Background

The Internet Shopping

Today the whole world is facing an "electronic" change affecting the way people communicate as well as transforming the entire value chain from producers and retailers to consumers. Traditional exchanges between individuals and firms have been revolutionized as Internet shopping is becoming a well-accepted way to purchase a variety of products and services. In 1998, over $22 billion in exchanges were made on the Internet, and the number is expected to reach $3.2 trillion in the year 2003 (Donthu and Garcia, 1999).

Many factors contribute to the development of the Internet market, some related to technological advances, some related the way the corporate world has changed its perceptions, and some related to changing lifestyles of consumers. The increasing number of companies that offer Internet access are providing consumers with a convenient and inexpensive way to become members of the Internet community. The development of better navigation software and search engines are making Internet visits a more pleasant and exciting experience. The increase in the quantity and quality of the available information on the Internet and the presence of well-known corporations and brands on the Internet are also generating higher interest among consumers. In addition, the development of secure systems that allow secure monetary transactions is accelerating Internet shopping.

The U.S. is by far one of the global leaders in the e-commerce arena. In 2000, among the U.S. respondents using the Internet, 74% have made at least one online purchase in the past twelve months and 87% expressed a desire to make an online purchase in the next year (Ernst & Young, 2001). The number of Internet purchases made by U.S. consumers averaged about 13. Three-quarters of online buyers reported an increase in the number of items purchased online. Online buyers in the U.S. spent, on average, $896 shopping on the Net, and three-quarters also increased the value of their online spending in the past 12 months. It is expected that both the number of online shoppers and the amount of their spending will continue to grow in the following years as the number of Internet users and the percentage of online buyers increase and the willingness of the Internet users to shop online increases.

The Internet Shopper

Shopping has become the fastest-growing use of the Internet, and almost 40 percent of the Internet users report shopping as a primary use of web. The total number of Internet shoppers has reached more than 20 million, and it is expected to continue growing. Even consumers who have not used the Internet to purchase goods and services claim to have used it for information searching that ultimately led to shopping in the traditional channels (Donthu, *et al.*, 1999).

While it is useful to profile the Internet *user*, from a marketing or advertising point of view it is more important to understand the Internet *shopper*. According to

Donthu and Garcia, a typical Internet *user* does not indicate a typical Internet *shopper*. They find out that online shoppers do not differ from non-shoppers on education and gender. However, Internet shoppers are older and earn more money than non-shoppers. It is because they are provided with a higher purchasing power and easier access to credit cards that made them to be more likely to purchase online. In addition, online shoppers value convenience are more willing to try new things and are less concerned about the risk involved.

However, online shoppers are less brand and price conscious than non-shopper. Since online shoppers belong to the higher income class, price becomes a less important benefit that Internet can provide them. The major purpose of their online shopping activity is to find satisfactory goods and services. Traditionally, consumers regard brand and price as symbols of the quality of products. But as more and more detailed product quality information is available on the Internet, brand and price are no longer a surrogate for product quality. It is also found that Internet shoppers spend more time browsing on the web sites to search for products and information than non-shoppers. Also they possess a more positive attitude toward Web advertising and direct marketing (Donthu, *et al.*, 1999).

Internet Shopping-Specialty Food and Beverage

With online shopping becoming a very popular activity among Internet users, e-commerce industry is growing. Now Internet shoppers are able to buy a great variety of products and services online, including books, clothing, music, flowers, grocery items, computer hardware, travel arrangement, investment products and banking services.

Specialty food and beverage producers also have hopes for Internet commerce. Although online specialty food and beverage shopping comprises only a small part of online shopping, it shares many characteristics with the general Internet shopping.

The National Retail Federation (NRF) estimates that online food and beverage spending comprised $130 million in July, 2000 (NRF, 2000). The profitability indicates that the Internet has become a more and more attractive marketplace for specialty food and beverage retailers even though online sales only constitute a small portion of the total retail grocery sales. Therefore, it is not surprising to notice the development of online food and beverage sales in recent years. Specialty food and beverage producers are coming online every day. Their presence has grown from a total of merely over 100 firms in October 1995 to nearly 6000 food and drink companies, listed by the Yahoo Internet search engine, in 2001. This phenomenon strengthens the fact that specialty food and drink companies are increasingly pursuing this emerging marketing option -- online sales.

Incentives for This Research

By understanding what factors have substantial impacts on consumer's decisions about online products, specialty food and drink companies can strengthen their ability to develop effective marketing strategies, and as a result, improve their sales income. Since online specialty food and beverage sales have only recently developed, research on this newly developed subject has just started and is far from fully understanding this new market. Previous work concentrated on examining the demographic profile of the Internet users who visit food-related sites and identifying subsets of this population most likely to conduct online transactions. While this has helped identify some important characteristics of online specialty food and beverage buyers, questions like why Internet users with such demographic characteristics have a greater likelihood of shopping online, what motivates them to behave this way and to what extent these motivators influence their purchasing decisions need further analysis.

To date, little research that examines the motivating factors behind online shopping has taken place. Earlier studies on this aspect have led to conflicting information. Some have reported price to be the deterministic factor (Food Marketing Institute, Ernst & Young, 1999). Whereas others find that price is less important than other benefits derived from shopping on the Internet (Donthu *et al.*, 1999). This study tries to investigate food-related Web sites visitors from a life-style characteristic perspective and identify the particular life-style characteristics that the visitors with a higher probability of shopping online possess. Life-style characteristics depict a person's values, attitudes toward social phenomenon and his/her own personality traits. All of these, no matter how formed can, to some extent, portray online purchasers or potential buyers' behavior. For example, by adopting e-commerce, consumers often perceive some risk in the quality of information on the seller websites and Internet transactions. If the consumer is an extremely risk-averse person, the probability of his online purchasing activity will be rather low. Besides, since online shopping has only come about in less than 10 years, conservative consumers may still have suspicions toward it, therefore insist on traditional grocery shopping. Whereas innovative shoppers, who would like to engage in different ways of in-home shopping and are more willing to try new things, will have more interests in this newly developed marketing option and their probability of purchasing online will be correspondingly high.

To fully explaining consumers' online purchasing behavior, a study on identifying the determinants of consumers' behavior in an e-commerce environment is needed. It is believed that through an investigation of consumers' life-style characteristics combined with their demographic characteristics, factors that influence consumers' online purchasing decisions will be clearer. An online survey with the intention of gathering information on respondents life-style characteristics as well as demographic characteristics was posted as an Internet document on the University of Maine server from September to October, 2001. The resulting data set from this online survey was used for this study.

Another issue that deserves attention is that the previous studies are not predictive: they only give a general description of the likelihood of purchasing specialty food and beverage online for Internet users with certain kinds of demographic characteristics, while little information is provided on an individual basis. This problem will also try to be solved in this research. By acknowledging an individual's general personality trait, such as whether he or she is a fashion follower; whether he or she likes to be a leader, we can have an idea of the extent this particular individual would like to shop online.

Objectives

Corresponding to the development of the Internet commerce, there is an obvious increase in the revenue and profitability of some of the major Web-based businesses. In order to keep such kind of growth for large businesses whereas encourage middle- and small-sized businesses to take advantage of the growth, the e-commerce industry needs to better understand their audiences. The development of online specialty food and beverage sales and the incompleteness of previous research intrigue further studies concentrated on better understanding e-commerce industry and consumers' behavior in this environment. The primary goal of this study is to make further analysis on the determinants of online specialty food and beverage consumers' purchasing decision and develop expanded models that can better predict an individual's probability of shopping online. Based on demographic characteristics and life-style characteristics, this research will examine the electronic market for food and beverage products and explain consumers' choices for these products. Therefore, it is intended improved information for companies marketing food and beverage products online and strengthen their abilities to gather more accurate marketing intelligence and adopt effective market strategies. Specifically, the objective of this study is to identify the determinants of consumer's behavior and choice for online purchasing decisions.

To meet this objective, a factor analysis and cluster analysis of consumer behavior in an e-commerce environment are developed and estimated. The findings from this study can help decision makers of online specialty food and beverage companies to better understand what motivates their consumers to buy online and identify those who can be their potential buyers. It will provide a basis for developing what strategies could be taken by online specialty food and beverage companies to enlarge their target market as well as to encourage more spending.

Organization

Chapter II presents a descriptive analysis of online specialty food and beverage marketing. Specifically, it provides information on the commodities being exchanged in this special market, as well as the demographic profile of this particular section of online consumers. An initial description of life-style characteristics will be presented.

Chapter III reviews previous studies on consumer behavior models for market segmentation as well as the models for investigating consumer behavior in an e-commerce environment. Literature on factor and cluster analyses is presented along with an extensive discussion as to how three studies have used these techniques to segment the U.S. elderly market based on their lifestyle characteristics.

Chapter IV presents the development and specification of the models of factor and cluster analyses for this study. The rationale for factor and cluster analyses is also discussed.

The estimation method and estimation results are given in Chapter VI. A summary of the major findings of the research is reported in Chapter VII. Limitations of this study are discussed and recommendations for future efforts are presented.

CHAPTER II

AN OVERVIEW OF SPECIALTY FOOD INDUSTRY

Specialty Food Retailing Industry

Development of Specialty Food Industry

Although slower population growth, lifestyle changes, and the aging of the population have resulted in a flat trend of sales in the U.S. retail food stores, these factors have generated substantial variations in demand for certain types of food. Specifically, the demand for specialty food is exceptionally strong, as aging, health conscious, convenience oriented Americans develop more sophisticated palates and place greater emphasis on food quality. Nowadays, specialty food sales are experiencing an obvious growth and becoming more and more important to grocery shoppers. In the middle 1990's, about 81% of consumers in the U.S. report buying specialty food items at least once every six months. Nearly 20% of the household in the U.S. can be classified as medium to heavy consumers of specialty foods (Dietrich, 1992). The increased demand for specialty foods dates back to the middle 1980's when the industry experienced remarkable annual growth rates of 15-20%. More recently, the total market value of specialty foods in North America has been estimated at 20-30 billion U.S. dollars per year with an expected growth rate of 10% per year throughout the decade (Kezis, *et al.*, 1997).

As demand has grown, the number of new specialty foods introduced each year continues to climb, along with the number of retail outlets offering these items. Most supermarkets now have an expanded line of specialty foods, both packaged and fresh. Although consumers today still buy most of their specialty foods at the supermarket, they are increasingly likely to turn to the specialty food store and the Internet for a wide variety of particular products.

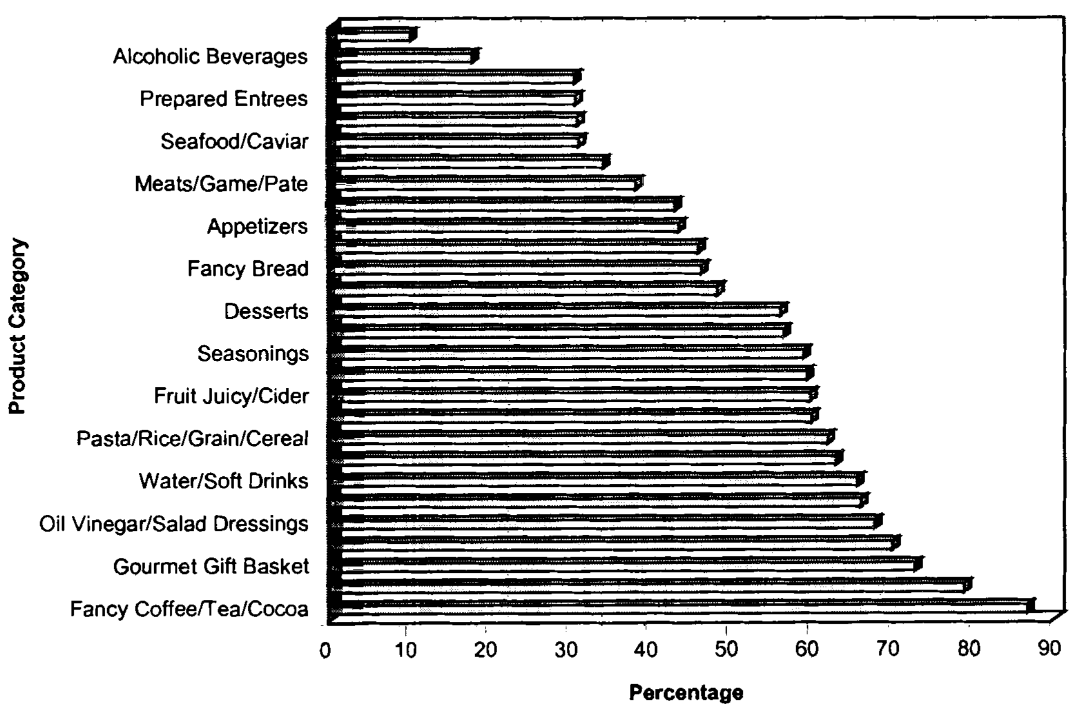
Specialty Food Products

There is no consensus on the definition of specialty food, but the category generally refers to "value-added, premium priced items that are distinguished in terms of one or more characteristics, such as the quality of ingredients, sensory appeal, origin (regional or ethnic), presentation (branding or packaging) or product formulation" (Kezis, *et al.*, 1997). Furthermore, hard-to-find items that are not available at local grocery stores can also be categorized into specialty foods.

According to a recent survey, gift buying is the most often mentioned reason for customers to patronize specialty food stores or departments. This phenomenon also corresponds to the fact that sales patterns of specialty food for Spring, Summer and Winter are similar, with an average of nearly 20% of the total sales during any of these seasons whereas the average sales volume doubles to 42% during the Fall and holiday shopping season. Although quality, uniqueness, packaging, healthfulness of product and convenience are the factors often considered by customers when selecting gifts from among various alternatives, product uniqueness and assured quality are the most important factors influencing specialty food consumers to make their purchasing decisions (Kezis, *et al.*, 1997).

Figure 2.1 presents a list of food product categories often carried by specialty food retailers in order from the most to the least. As can be seen from the figure, Fancy Coffee/Tea/Cocoa was the most popular product line followed by Candy/Chocolate and Gourmet Gift Basket. All of these specialty food categories were carried by more than 70% of retailers. The high prevalence may be due to their relatively longer shelf life. On the contrary, Seafood, Ice Cream and Fruits/Vegetables, which are perishable products and have limited shelf life, were carried only by less than one third of the retailers (Kezis, *et al.*, 1997).

Figure 2.1: Specialty Food Product Categories Carried by Retailers



(Source: Kezis, *et al.*, 1997.)

Specialty Food Retailers

According to the study on the profile of the specialty food retailing industry in the eastern U.S., specialty food retailers are extremely diverse so that it is difficult to portray the profile of a typical specialty food retailer. It is shown that specialty food retail operations are located in a remarkably similar proportion in shopping malls, downtown streets and freestanding buildings. The types of specialty food businesses range from those exclusively selling specialty food items, to gift basket operations in which most of the business is conducted through mail, and to convenience store where specialty foods are carried only as a small section. Specialty food stores constitute 38.2% of the group, taking up the largest proportion. Among stores with mixed product offerings, the percentile of sales from specialty foods is bi-modal, with the heavier loadings on the extreme ends. Thirty-two percent of the retailers are less than 25% of total sales, whereas another 36% claimed more than 75% of their sales are from specialty food (Kesiz, *et al.*, 1997).

Thus substantial variations exist in types and size of specialty food retailers. Small start-up operations and large supermarket stores both have the opportunity to sell specialty foods. Generally speaking, specialty food stores, which focus exclusively on specialty food, are relatively small in sales volume. They eagerly seek chances to reach broader, potentially more profitable markets. Therefore, as the use of World Wide Web (WWW) expands, these small specialty food businesses may also have hopes for Internet commerce.

Online Specialty Food and Beverage Marketing

Development of Online Grocery Shopping

Current lifestyles have led to great changes in American grocery shopping habits. With the increasing numbers of single-individual and dual-income households, as well as a growing need for leisure time, consumers' tolerance for doing routine daily tasks such as grocery shopping has been reduced. The time that these families would like to spend in the grocery store becomes less and less. On the other hand, computers and the Internet are commonplace as the result of enormous advances in technology (especially in IT[1] sphere). According to International Data Corporation (IDC), 53% of the households in the United States have PCs, and 41% of all households are connected to the Internet. About 17% of the population has shopped online from home, a percentage of whom are grocery shoppers.

It is always true that businesses that can respond to the changing needs of consumers are most likely to prosper. With the changing life-style and the corresponding various expectations toward grocery stores, online grocery shopping becomes a technological solution that can assist retailers to meet these needs and thus can be an effective tool for compensating the perceived disadvantages of traditional grocery shopping.

Online grocery shoppers seem to be satisfied with their experience of buying foods or drinks on the Internet. According to the recent FMI online shopper survey, more than four-fifths of all shoppers reported their orders were fresh, correct, intact and timely. The majority of shoppers rated their overall online experience as excellent. In addition,

[1] IT is the abbreviation of the term Information Technology.

more than half of the all shoppers are very likely to purchase groceries on the Internet again.

Food is the most popular grocery item bought online. Eight-eight percent of the respondents reported that they have purchased food online. Fifty-three percent have bought health and beauty-care items, 43% interested in vitamins, herbal remedies and diet aids. Among the available online food items, 63% of the respondents claimed that they have purchased non-perishable online, the most popular food items; about 33% stated that they had bought frozen foods or meat online; whereas less than 25% have bought produce by means of the Internet (The e-tail Experience, 2000). This phenomenon indicates that perishable food products are less demanded by online shoppers than non-perishable or those with a longer shelf life.

Although online shopping provides consumers with convenience and efficiency, grocery shoppers still hold different views toward online groceries. Only 10% of the overall grocery shoppers have ever purchased online. Furthermore, most of them made their online purchase from an online-only grocer. This was due to the lack of availability of the product from primary grocery stores. This fact, to some extent, hinders the development of online grocery shopping. For those who only trust what they have bought by personal touch, online shopping may be an inappropriate vehicle for commerce. In addition, inconvenient delivery times and methods can be another obstacle.

Online specialty food and beverage shopping only comprise a small proportion of the total online grocery shopping, however, specialty food shopping shares many common characteristics with overall grocery shopping. Therefore, the analysis of online grocery shopping may provide useful insight into studies on online specialty food shopping, which will be introduced in the following.

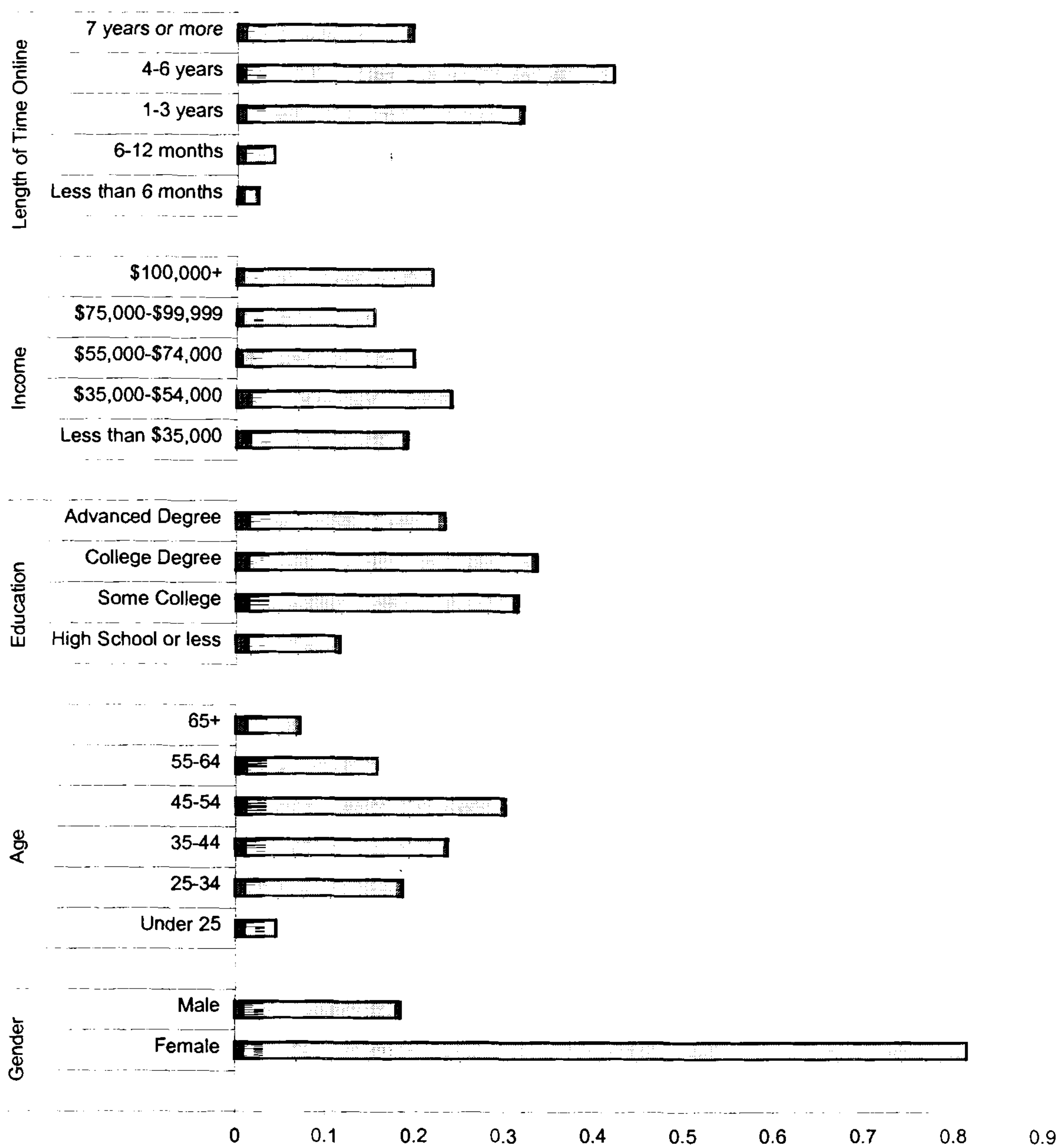
Online Specialty Food and Beverage Consumers

The demographic profile of Internet users who visit food and drink sites online differs from the Internet population as a whole. They are primarily female, middle-aged, wealthy and well-educated (White, 2001). However this population shares many characteristics with traditional retail specialty food shoppers.

Since 1996, White has conducted a series of surveys investigating the demographic characteristics of Internet visitors to food and beverage sites and their online shopping activities. Results indicate that Internet users who visit food and drink sites differ from both the general Internet population and the U.S. population at large (White, 1999). As shown in his most recent online specialty food consumer survey, there is an obvious tendency toward gender disparity among the food-site visitors. The gender of visitors to food and beverage sites continues to differ from the Internet population as a whole with 81.5% female representation (White, 2001). This has been statistically unchanged from 1999.

Same as in 2000, the results from 2001 survey shows that visitors to food and drink sites were not young as the majority (53.8%) aged between 35 to 54. With the more and more widespread of Internet use, the educational level of its users is decreasing. This is also true for specialty-food-site visitors. In 1996, nearly two-thirds of the visitors held at least a college degree. While in 2001, only a little bit over 50% of the respondents reported that they had a college degree and 11.6% stated that they had not received any education beyond high school. But, generally speaking, Internet visitors to food and drink sites are still in a higher education level compared to the Internet population. Like education, respondents reported their household income higher than national average. More than one-half of the Internet food-site visitors have an average household income above $55,000. In addition, 21.8% of them stated that their income is over $100,000. A long history of Internet use also distinguishes Internet food-related site visitors from the general Internet population. Nearly sixty-two percent of the survey respondents reported using the Internet for more than four years, while less than 10% reported using the Internet for no more than a year. These demographic characteristics of food related site visitors have been shown in Figure 2.2 (White, 2000 and 2001).

Figure 2.2: Demographic Profile of Internet Users Who Visit Food and Beverage Sites



(Source: White and Manning, 2001.)

Among Internet food-site visitors, 85% stated that they have bought at least one specialty food item online in the past six months though their purchasing activities are infrequent (White, 2001). This indicates that the majority of the food-site visitors are online specialty food shoppers. As a result, studies on food-site visitors provide a solid basis for more specific investigation on online specialty food consumers. As previous studies primarily focused on the demographic profile of the food-site visitors, they provided an initial impression of the demographic characteristics of online specialty food buyers.

The same items that specialty food retailers most often carry in their product line, e.g. Fancy Coffee/Tea/Cocoa, are also the product categories most frequently purchased by online specialty food consumers. Nearly 40% of online buyers reported that they had made at least one purchase of this product category in the previous six months. Ice Cream/Frozen Desserts and Dessert Topping are the least frequent purchased items with less than 1.5% online consumers indicating having made such a purchase (White, 2000).

With regard to the sources of consumers' online purchase, the respondent population identified approximately 533 different companies and their respective weight of online sales has been shown in Table 2.1. Nine companies including Omaha Steaks, Penzeys Spices, Harry and David, Gevalia, Tavolo, King Arthur Flour, Wine.com, Dean & Deluca, and Williams Sonoma, were the major sources of online shoppers'purchases and represented nearly one-third of the overall online purchase. Five additional companies including Cooking.com, Ethnic Grocer, eVineyard, Wolferman's and Peets Coffee accounted for additional 7.4% of all purchases. The remaining 60.3% were distributed across the other 519 companies (White, 2001).

Table 2.1: Sources of Online Specialty Food Buyers' Most Recent Purchase

Sources	Percent
Omaha Steaks	5.90%
Penzeys Spices	5.20%
Harry and David	4.70%
Gevalia	3.30%
Tavolo	3.00%
King Arthur Flour	2.80%
Wine.com	2.80%
Dean & Deluca	2.60%
Williams Sonoma	2.20%
Cooking.com	1.80%
Ethnic Grocer	1.80%
eVineyard	1.30%
Wolferman's	1.30%
Peets Coffee	1.20%
519 Other Stores	60.30%

(Source: White and Manning, 2001)

CHAPTER 3

LITERATURE REVIEW

Consumer behavior is one of the most important, dynamic scientific fields in market research. With consumer behavior analysis, many important marketing statistics, such as market share, a brand's purchase probability and so on, can be evaluated by building formal models of consumer behavior. However, consumers differ from each other in a range of characteristics like preference, income, gender, personality, etc., which implies that it is impossible to employ a single model to incorporate all of these variables. Therefore, a variety of models are preferred to adequately differentiate one consumer's behavior from another.

In the last forty years, consumer behavior models have been substantially developed. These models, contrary to verbal models, are explicitly with a mathematical structure. With their rapid speed of proliferation, it is not surprising to notice the models exist in a bewildering array of forms. Roughly speaking, formal models of consumer behavior may be classified into four broad types, including information-processing models, experimental and other linear models, large-system models, and stochastic models (Bettman and Jones, 1972).

Preview of the Formal Consumer Behavior Models[2]

Information-processing Models

The most obvious characteristic that distinguishes information-process models from other consumer behavior models comes from their basic assumption, which states that every individual continually receives information from his environment and actively processes this information when making choices. This kind of model concentrates on a particular individual by using protocol data[3]. As a result, it is highly idiosyncratic.

Much of the early work in this field was done in psychology and computer science. Particular information-processing models with regard to consumer choice are few. Alexis *et al.* (1968) modeled the case of women's clothing decisions. Bettman (1970) modeled consumers' choices for grocery products, and Russ (1971) modeled the choices of several subjects for small durable goods in a laboratory setting.

Experimental and Other Linear Models

Models in this classification have diverse content but have a common formal mathematical structure – linear structure. Generally speaking, the model might be portrayed as

$$f(y) = \sum_i g_i(x_i) + \varepsilon$$

[2] The introduction of the consumer behavior models in this section is based on Bettman and Jones, 1972.

[3] The protocol data is obtained by having each subject think out loud while he is performing the behavior being modeled.

where y is the special aspect being modeled and x_i's are the explanatory factors. The stochastic error term ε is a random element and contains those factors that influence the dependent variable y but are not explicitly expressed in the formulation. Another special feature of linear models is that most of them are designed to describe the market rather than an individual consumer.

Large-system Models

Large-system models are characterized by a somewhat simplified formal model fitting within a broad general structure of postulated interrelationships. Since these models are generated corresponding to the reductions of comprehensive verbal description of consumer behavior, a lot of mathematical formulations are employed. Mathematical diversity within these types of models is much more than others. There are three major models of this type: Amstutz's microanalytic simulation consumer-behavior model, Farley and Ring's linear realization of the Howard-Sheth model, and Nocosia's differential-equation model, respectively.

Stochastic Models

Stochastic models have been used extensively in analyzing consumer behavior. The use of stochastic models to study brand choice as well as purchase incidence was introduced by Kuehn (1958, 1962). He proposed "purchase event feedback" as a potential explanatory mechanism for brand choice. According to Kuehn, consumers who have purchased a certain brand for several times in the past might have a much higher propensity to choose that brand on the next purchase occasion. This can be labeled as "purchase event feedback", and it became the only influential factor for explaining consumer's choice behavior in his linear learning model.

Frank (1962) put forward an alternative method for describing Kuehn's data. He assumed the existence of heterogeneity, a characteristic delineating different consumers who have different initial attitudes toward a special brand. Constant attitude of each consumer toward this special brand over time and purchase events is another basic assumption of Frank's model. However, a great divergence appears at this point. Whether purchase event feedback or heterogeneity was the controlling factor of the consumer behavior that Kuehn had observed became the focus of much of the research in this field in the following ten years. Massy *et al.* (1970) summarized the results from previous research and concluded that both are influential.

More recent stochastic models accepted the tradition of purchase event feedback and heterogeneity developed earlier and tried to extend the models to incorporate additional explanatory variables. Unfortunately, these models were idiosyncratic in the sense that they were unable to include any explanatory variables except for those that were specifically designed in the model. This phenomenon lasted until a new brand choice model was proposed by Jones and Zufryden (1980,1982). This model combines heterogeneity, purchase incidence and logit regression concepts.

From a conceptual perspective, a stochastic model can be divided into two parts: an individual behavior model and the aggregation mechanism. The individual model depicts a special aspect of an individual consumer's purchase behavior and is represented by an explicit function of the major determinants. The probability of purchase is what to modeled. The aggregation mechanism has two forms. Assuming every individual is the same in all aspects of behavior is the most straightforward one. However, many of the current models took a second form which assumed that different individuals react somewhat differently to the same set of stimuli. This is a more accurate description of the real world but the aggregation problem becomes more complicated in the second case.

For the current study, disaggregate attribute choice models are more appropriate to analyze factors affecting consumer's choice of purchasing online. As opposite to the aggregate model, which employs generalized data concerning to the population such as mean, a disaggregate model uses individual observation to estimate parameters of the population. Disaggregate attribute choice models attempt to provide not only the aggregate choice prediction but also diagnostic information about the function of various attributes in the choice process.

The development of this kind of choice model in marketing research began in the 1960's. But from the initial models put forward by Rosenberg (1956) and Fishbein (1967) to the current methods of linear regression, discriminant analysis, logit and probit, all are generally estimated under the assumption that all members of the population are homogeneous in their choice function. Unfortunately, this assumption runs counter to one of the basic concepts of marketing theory --- market segmentation. This perspective, that consumer population can be segmented based on different preferences and perceptions, has been recognized by practitioners for a long time. Development and research results in this field are presented in the next part.

Consumer Behavior Models for Market Segmentation Studies

Market segmentation[4] analysis in consumer behavior research has been applied to various product categories and the consumer groups defined through various methods. Wildt and McCann (1980) investigated market segmentation from the explanation of the variance in observed consumption behavior perspective. Previously, normal regression models were employed in this kind of analysis and the ordinary least squares (OLS) has the desirable properties of unbiasedness and minimum variance in estimating the unknown coefficient (BLUE) (Griffth, *et al.*, 1993).

But the appropriateness of using OLS in analyzing consumption data has been questioned. Ehrenberg (1972) proposed that a Poisson process was better in representing typical measures of consumption and purchase. Morrison (1973) further argued that the usual R^2 statistic obtained from an OLS regression is not appropriate to evaluate the results of the usual segmentation study. Wildt *et al.* (1980) recognized the inherent component of randomness in the variation of observed consumption behavior and expanded the preceding formulation of purchase process. In the new model, the disturbance term consist of not only the difference between the true mean purchase rate and the predicted mean purchase rate but also the difference between the true mean purchase rate and the number of units purchased in a given time period. This model, labeled as iterative generalized least squares (IGIS), was used to compare two OLS models, the first with the same criterion and prediction variables and a second with some changes in the criterion variable. The result showed that the IGLS model offers a more realistic representation of the process and is more consistent with available empirical

---

[4] Market segmentation is to divide the total market into smaller, relatively homogeneous segments.

evidence. This study, aiming at separating the explanation of individual purchases from the explanation of the mean purchase rate, provides insight into the need for research in the latter field. This result became strong evidence in refuting the homogeneous assumption of stochastic models and initiated further analysis in marketing segmentation.

Jones and Zufryden (1980,1982) proposed a new stochastic choice model that integrates brand choice and purchase incidence behavior. The final results indicated that the consumer population could be segmented based on demographic characteristics. The overall model is comprised of two components. A logit model is put forward to estimate the probability of selecting a certain brand given a function of various explanatory variables, which can be regarded as an expansion of previous choice models. A weighted least squares (WLS) estimation procedure is developed for the logit model. The other component is a Poisson model used to describe the purchase incidence behavior for the product category. Both components take into account the heterogeneity of individual consumers within a general population.

The data for a frequently purchased, non-durable good were used to test the new model, and the selected certain brand is one of the market leaders, taking approximately 8% of the whole market share. According to Jones and Zufryden, the presence of children in the household, income, and relative price are the influential factors in explaining brand purchase behavior. Of these, the relative net price of the special brand appears to be a critical variable considering its magnitude and the significance of the corresponding parameter coefficient. In addition, the identified important demographic characteristics (income, presence of children, etc.) also suggest market segmentation opportunities. Thus the study indicates that demographic variables can distinguish among groups when the choice is predicted at the individual level, which is contrary to the conclusion of homogeneity from previous research.

Though it has been long realized that a potential consumer population may contain segments with very different preference toward a certain product class, with the lack of additional information to reduce heterogeneity, this problem has always been neglected in the early research. Wilkie and Pessemier (1973) reviewed forty-two empirical articles using a single-function model. Only two of these considered the possible existence of market segmentation with different perceptions and preferences. Malhotra (1984) reviewed the logit application in marketing literature, and none of the studies directly stated the segmentation issues.

Historically, the testing of homogeneity for a disaggregate choice model has been a clumsy procedure: first a preliminary knowledge of the segments is required and then submitted to a multiple-function model. Two directions have been developed in the empirical applications. In decompositional models such as conjoint analysis, the individual-level utility function has been estimated. Subsequently, the segments of individuals have been formed by aggregating those with similar utility functions. In compositional models such as logit, the data was collected based on a priori knowledge of segments, which enabled the researchers to obtain information on the explanatory variable in a manner that reduces the heterogeneity on a particular dimension.

The log-likelihood test used by Gensch (1985) makes testing for segments fairly easy and straightforward with the logit algorithm. The overall goodness of fit criterion is the value of the log-likelihood function evaluated at the estimated coefficients. Given the assumption of independence among individuals, the values for each possible segment in the sample may be added together. If the sum value is larger than that of the one-logit fit for the entire sample, it indicates that particular segmentation scheme fits the data better and the assumption of homogeneity with regard to the attributes values is unacceptable.

Gensch used the data from a Santa Monica Freeway survey in an empirical analysis. First nineteen demographic variables for each observation were submitted to a cluster analysis which resulted in three distinct clusters. Since the additive value of log-likelihood functions of each segment was significant at the 5% level, the null hypothesis of homogeneous logit function was rejected. An effective analysis suggested various useful strategies in persuading auto drivers to switch to bus transportation. Gensch finally concluded that meaningful segmentation could lead to more accurate results of choice models and, more importantly, result in effective managerial strategies for different segments.

Recently, more attention in marketing literature has been turned to consumer segmentation and numerous bases for this useful concept have been advanced. Grover and Srinivasan (1987) proposed that similar brand choice probabilities, a basis for behavioral segmentation, also generate patterns of brand switching that can be used to define the product market structure. Shugan (1987) demonstrated that temporal changes in brand prices could be used to calibrate the size of various segments as well as the location of brands in the market structure. Consumer segmentation, market structure and price sensitivity are thus combined together.

Kamakura and Russell (1989) suggested a new segmentation approach to identify the underlying determinants of brand switching behavior and aggregate response to price changes using household-level data. This model rested on the assumption that consumers can be segmented into different groups characterized by similar preference and a single price sensitivity parameter. In deriving the model, random utility theory was the initial step. The random utility of consumer k to purchase brand j at the time period t is

$$U_{jkt} = u_{jk} + \beta_k X_{jkt} + \varepsilon_{jkt} \tag{3.1}$$

where u_{jk} is the intrinsic utility consumer k obtained from brand j, β_k is the price parameter for consumer k, X_{jkt} is the net price of brand j that is available to consumer k at time t and ε_{jkt} is the random disturbance term.

A multinomial logit model is then proposed to represent the conditional probability of selecting brand j for consumer k at purchase occasion t,

$$P_j(u_k, \beta_k, X_{kt}) = \exp(u_{jk} + \beta_k X_{jkt}) / \sum_{j'} \exp(u_{j'k} + \beta_k X_{j'kt}) \tag{3.2}$$

The key idea of this model is to use the choice probability corresponding to the various segments to express an individual consumer's choice probability. This major procedure is obtained by re-labeling equation 3.2, and, therefore, the probability of choosing brand j, conditional on consumer k being a member of segment i is,

$$P_j(u_i, \beta_i, X_{kt}) = \exp(u_{ji} + \beta_i X_{jkt}) / \sum_{j'} \exp(u_{j'i} + \beta_i X_{j'kt}) \tag{3.3}$$

Maximum likelihood procedure is applied to equation 3.3 to get the estimates of the model parameters conditional on the assumed number of switching segments. To specify the number of segments, Akaike's information criterion[5] is employed,

$$AIC = -2(LL - p) / N \tag{3.4}$$

---

[5] Akaike's information criterion (AIC) is given by, $AIC = \log(V) + \frac{2d}{N}$, where V is the loss function, d is the number of estimated parameters, and N is the number of estimation data.

where LL is the maximum value of the log likelihood, p is the number of estimated coefficients and N is the total number of observations.

Kamakura and Russell applied the data for a median frequently purchased food item into this new probabilistic model, classified five meaningful segments and developed a representation of market structure in terms of both brand preference and price elasticities.

Consumer Behavior in an E-commerce Environment

The development and prosperity of electronic commerce has been increasing rapidly due to the effectiveness and convenience it can bring to the consumers. E-commerce is fundamentally changing the way consumers purchase goods and services, and, the focus of many marketing studies. Pervious research focusing on electronic commerce and segmentation analysis are two important areas for this study,

It is projected that the Internet will generate consumer and business-to-business sales in excess of $294 billion by 2002 (Computer World, 1998). In order to increase market shares, develop reasonable marketing strategies and thus better promote products in this promising online market, marketers should understand what factors influence online purchasing behavior and the extent to which they do.

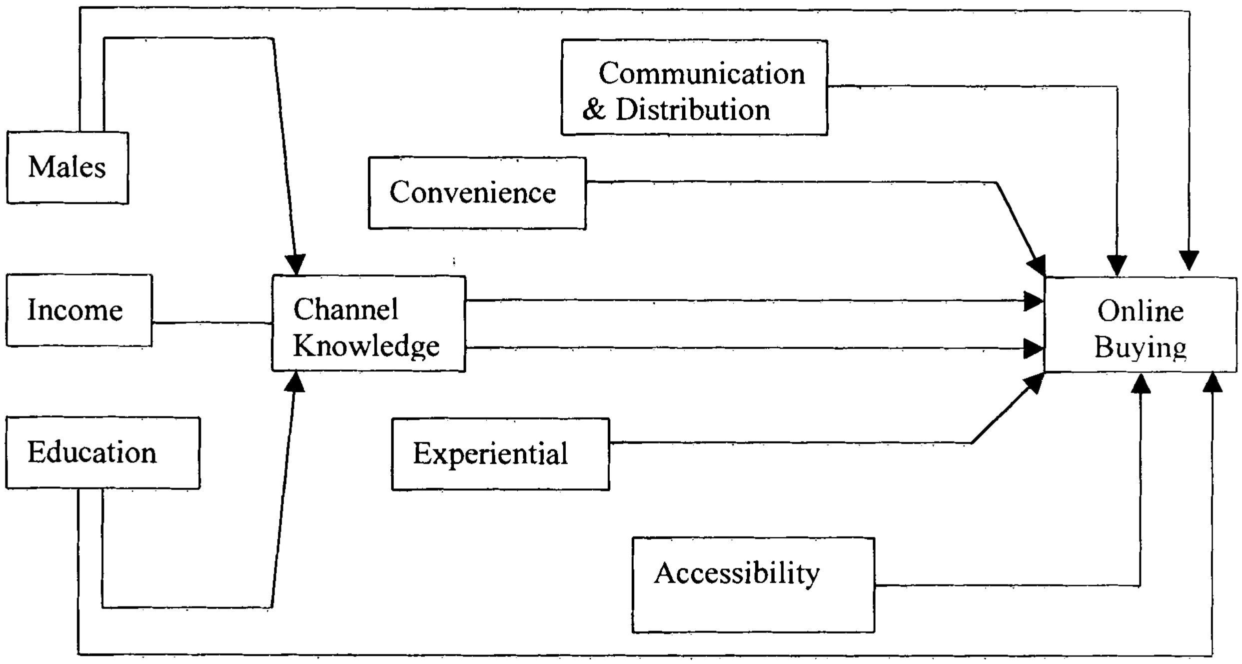
Bellman *et al.* (1999) conducted a survey to address the predictors of online buying behavior. Besides consumer demographics, lifestyle characteristics have also been used to explain why people purchase online and the amount of money they spend there. A survey questionnaire with 62 questions about potential respondents' online behavior, attitudes towards Internet communication and privacy issues, as well as standard demographic questions was posted on the web in October 1997. More than 10,000 people completed this survey, providing the data on which the research depended.

Bellman *et al.* used logistic regression to predict actual purchases by splitting the large sample randomly into two samples for analysis. A calibration sample used a stepwise procedure to add candidate variables into the regression. The final regression for this calibration sample predicted whether members of the second sample (holdout sample) would or would not purchase online. The process was repeated to calibrate the regression using the original holdout sample and cross-validate this regression equation in the original calibration sample. And they concluded that looking at product information on the Internet is the factor with the highest-value estimate --- approximately 0.23, thus is the most important predictor of online buying behavior. This can be further demonstrated by the following example: a typical online buyer has been online for years, received lots of emails everyday and used the Internet every week. They used the Internet for many of their everyday activities, therefore it is quite natural for them to search for product information online, and the more information he obtain, the higher probability of their purchasing online. Time limitation is another factor influencing a person's decision to shop online. That is because these individuals have less time to search for and buy products and services in traditional ways. Thus e-commerce is a faster and more convenient way compared to physically visiting shopping stores caters to their needs.

Li *et al.* (1999) attempted to identify the factors that could be used to predict a customer's online buying behavior by testing a linear regression model. Their data were collected by ClickinResearch company from an online survey of national Internet users. The data were then cross-validated with other similar national surveys (i.e., Greenfield Online) before being used to test the model. Ten hypotheses which are related to marketing channels, shopping orientations, and consumer demographics are put forward before introducing the model. These hypotheses can be divided into three parts. For the channel theory part, researchers tried to figure out whether online purchasers perceive the Internet offers higher channel utilities and consider themselves more knowledgeable about the Web as a channel than non-purchasers. The difference is tested for frequent purchasers vs. occasional purchasers. For the shopping orientation part, which quantified customers' lifestyle factors, they tried to prove that online purchasers are more convenient-oriented, less experiential-oriented, and indifferent to recreation and price orientation. Whether similar situation applies to frequent purchasers and occasional purchasers is also to be tested. For the demographic part, they tried to show that online purchasers are better educated and with higher income than non-purchasers. Age differences are assumed to be insignificant among three types of online buying status as part of the hypothesis.

A conceptual framework that provides a better insight on factors that affect online purchase behavior and also a summarization of all the hypotheses is shown in Figure 3.1.

Figure 3.1: A Conceptual Model for Online Consumer Purchasing Behavior



(Source: Li, *et al.*, 1999)

The ten hypotheses were first tested with one-way ANOVA. As a result, most were fully supported and only a few partially supported. The two unsupported hypotheses included the assumed difference between occasional purchasers and frequent purchasers in convenience orientation and the gender difference between online purchasers in general and non-purchasers. Then multiple regression models depicting various factors and testing the extent these factors influence consumer buying behavior are built step by step in stages. Initially four demographics factors are included (model 1). On the basis of the first step, four shopping orientation factors are added (model 2). Channel knowledge factors are included as additional explanatory variables to model 2 (model 3), and three kinds of perceived channel utilities are also added (model 4). The results showed that model 1 can explain only 4.3% of the variance in online purchasing behavior. This percentage increased to 15.6% by including all the variables in model 2 and to 26.9% for model 3. The final model can predict 28.9% of online consumer purchasing behavior, which could be considered a moderately strong model. In the final model, only six variables are significant predictors, and they are education, convenience and experiential shopping orientations, channel knowledge, channel distribution utility, and channel accessibility.

Smith *et al.* (2000) used data from Internet shopbots, which are web-based services that provide "one-click" access to price and product information from various competing retailers, to investigate what factors customers value when making choices of buying books online. The data were collected from EvenBetter during the period of August 25 to November 1, 1999. From the initial statistical analysis, two important facts have been observed. These are that a high level of price dispersion exists among homogeneous books in the offer sets and consumers demonstrate a willingness to bypass the lowest cost retailers. In 2001, they continued their research, trying to use multinomial logit and nested logit models to systematically analyze the possibility of customer's willingness to pay a premium for a homogeneous product. Differences were generated from retailer branding and service quality. They found that shopbot customers care about the brand of the book retailer and specifically, they strongly prefer offers from such well-known retailers as Amazon.com. They further suggested that the reason for this phenomenon is consumers' use of brand name as a signal of reliability in service for non-contractible aspects of the product bundle such as shipping. Since books are a well-specified and homogeneous commodity, the fact that branding is important when customers select them from a number of online retailers implies that it will be more important in Internet markets for less homogeneous goods and services, especially when they have important non-contractible characteristics.

Lifestyle Characteristic Analysis in Consumer Behavior Models

This kind of analysis has been applied to the older American market. Elderly Americans have been regarded as invisible consumers and often ignored by marketers. However, recent recognition of the attractiveness of the mature market is based on the obvious increase in the size and growth rate of this segment of the population as well as their purchasing power. Demographic variables, such as age, marital status and socioeconomic status became the basis for early attempts to segment this market, but they resulted in only a small proportion of the variance being explained when predicting buyer behavior. French *et al.* (1985) used social gerontology theory to segment the elderly market, but this attempt has great limitations because of the nature of the data set.

In order to improve the model prediction, quantifying lifestyle differences existing among consumers has emerged as a viable solution. Sorce *et al.* (1989) extended the notion of variation in the mature market by segmenting this market on the basis of standard life-style characteristics. A life-style questionnaire was constructed to include 62 psychographic statements with the intention of measuring eight lifestyle dimensions. These dimensions included social activity; physical activity; financial behavior; attitudes toward family; physical security; perceived vulnerability to illness; risk-taking and self-reliance. The data from the psychographic statements were submitted to a principal components factor analysis with a varimax rotation. Under the selection criterion that the eigenvalue should be greater than one, five factors were obtained (Table 3.1), and 31% of the variance can be explained by these five factors.

Table 3.1: Principal Components Factor Analysis Result for Life-style Segmentation of the Mature Market

Factor	Eigenvalue	Example of Statement
Self-reliance	4.39	"I feel a lot of pride in using the things I've built or made by myself."
Social Venturesomeness	2.10	"I enjoy parties."
Security	1.91	"I avoid going out at night because I'm afraid I'll be mugged."
Physical activity	1.29	"I like to do something physically active at least once a day."
Family orientation	1.11	"I enjoy spending a lot of time with my family."

(Source: Sorce, *et al.*, 1989)

Next, the factor scores for each observation, which were a direct result from factor analysis, were applied to a cluster analysis. A squared Euclidean distance measure (centroid) and a furthest neighbor (complete linkage) clustering method were employed, and six clusters emerged. Sorce *et al.* also calculated the demographic profiles and the average factor scores for each cluster. The results showed that significant differences exist among the six revealing clusters on demographic variables and on all of the consumer attitudes and behaviors.

Although the effectiveness of segmenting the elderly market into distinct categories by lifestyle characteristics has been proved, little research has been done to analyze the impact of lifestyle segmentation on the purchasing behavior of older consumers. Shufeldt *et al.* (1996 and 1998) expanded the previous studies in a series of reports with two major purposes, one of which was to develop a lifestyle segmentation model of the mature market, and the other was to identify the relationships between the lifestyle segments and retail store attributes as well as factors influence purchasing behavior.

In 1996, Shufeldt *et al.* developed a questionnaire to determine the lifestyle characteristics and preferred retail store attributes of shoppers. Lifestyle characteristics were measured by the respondents' activities, interests, and opinions. Statements used to identify the lifestyles were based on a five-point scale. Similarly, the lifestyle data were also submitted to a factor analysis that performed a principal components analysis along with a varimax rotation. The extraction of factors obeyed the criterion of eigenvalues greater than one. Then the extracted factors were submitted to the varclus procedure, which revealed five clusters as proposed by Sorce *et al.* The chi-square test of homogeneity was used to test for equal variance of the clusters. The results showed that the groups were independent at 5% significance level. By applying the general linear framework that permits unequal cells in the ANOVA procedure, the resulted five clusters explained 16.8% of the variance at 5% level. Therefore, it can be stated that there are significant differences among the five groups. Variables describing store attributes were also submitted to a factor analysis. Then ANOVA tests were used to determine whether the lifestyle clusters were different from each other considering store attributes. In addition, appropriate *post hoc* multiple comparison tests were also performed.

Two years later, Shefeldt *et al.* conducted a similar study with the intention to investigate whether the lifestyles of the elderly and other factors influence the purchase of over-the-counter drugs. ANOVA analysis was also used to determine their relationships.

CHAPTER 4

MODEL SPECIFICATION AND DEVELOPMENT

This chapter presents the development and specification of factor analysis and cluster analysis to define segments based on lifestyle variations and identify what kind of consumers may have a higher propensity for purchasing specialty food online. Ten factors are extracted from forty-two statements that are used to identify an individual's lifestyle characteristics and the resulting factor scores are submitted to a cluster procedure. The mean value of the variables representing the major demographic characteristics as well as that of the related online activities of each cluster are calculated and compared. Whether the likelihood of shopping online for one group is statistically different from another will be measured by a probit model.

Direct estimation of all the factors that may affect an individual's online shopping activities by employing a probit model has led to ambiguous results. In the probit model, unlike the usual linear statistical models, the parameter value is not directly interpretable as the effect of a change in an explanatory variable on the mean, or expected value, of the dependent variable. The sign of the parameter estimates indicates the direction of the relationship between the explanatory variable and the probability, and the magnitude of the change in the probability cannot be obtained unless the magnitude of the individual's utility function is known, implying the extent of the explanatory variable's impact on the incidence probability is not available. Therefore, even if the probit model have shown goodness-of-fit statistical estimates, it can merely provide information of whether the factor has a positive or negative influence on the individual consumer's probability of shopping online whereas yielding little information with regard to which factor has relatively greater influence on the consumer's online shopping activities (Griffth, *et al.*, 1993). In addition, the major objective of this research is to investigate whether lifestyle characteristics are determinants of consumer's online purchasing behavior. However, the probit model may obscure the effect of lifestyle characteristics by incorporating all the variables that may be related to consumer's online shopping activities. In this study, we argue against a probit model because of the unfruitful estimation results it can provide. Instead, a cluster analysis is used because it not only directly relates lifestyle characteristics to an individual consumer's online shopping activities but also generates predictive results. In the following section, a factor analysis and a cluster analysis will be developed respectively in order to investigate the relationship between the respondents' lifestyle characteristics and their online purchasing behavior.

Details of the data to be used for model estimation and analysis are discussed in Chapter V. Chapter VI concentrates on the estimation results.

Factor Analysis

In order to get information from the widest possible variety of variables and increase the chances of hitting upon those that will expand our knowledge about life-style characteristics, forty-two statements regarding lifestyle characteristics were included in the survey. Thus, a higher-order data reduction technique, which can identify and summarize the many inter-relationships existing among the individual variables, is desirable. In this case, factor analysis satisfies this requirement. It consists of a family of statistical techniques concerned with the reduction of a set of correlated variables by using a smaller set of "derived" variables, or factors. Alternatively, we may think of factor analysis as removing the duplicated information from among a set of variables (Kachigan, 1991).

Common Factor Analysis

Factor analysis refers to all methods of data analysis using matrix factors, including common factor analysis and component analysis. Though the two techniques function very similarly and are used for the same purpose, they are quite different in terms of underlying assumptions and mathematics.

In common factor analysis, it is assumed that the variance of a single variable can be decomposed into common variance that is shared by other variables included in the model and unique variance that is unique to the particular variable and includes the error component. In other words, each input variable can be viewed as a weighted combination of factors in common factor analysis. However, only the common variance is the subject of the common factor analysis.

Thus, the equation for the common factor model is

$$y_{ij} = x_{i1}b_{1j} + x_{i2}b_{2j} + \ldots + x_{ik}b_{kj} + \ldots + x_{iq}b_{qj} + e_{ij} \tag{4.1}$$

where

y_{ij} is the value of the ith objects on the jth variable.

x_{ik} is the value of the ith objects on the kth common factor.

b_{kj} is the regression coefficient of the kth common factor for predicting the jth variable.

e_{ij} is the value of the ith objects on the jth unique factor.

q is the number of common factors.

Given the assumptions that the unique factors are uncorrelated both with each other and with the common factors, the correlations among the observed variables are explained by the factors.

The Centroid method is most often used for the extraction of factors in common factor analysis. The basic principles are presented by four main stages. Since the diagonal cells of a correlation matrix derived from the original data set are always empty, the first step involves making the best guess of communities, the proportions of common-factor variance and also the value of the diagonal cells. With the assumed communities, T represents the sum of all the elements in the matrix, and the first-factor loadings are given by the equation

$$a_1 = \frac{E_j}{\sqrt{T}} = mE_j \tag{4.2}$$

where

a_1 is the first centroid factor loading

E_j is the sum of correlations with the observed variable j

T is sum of all coefficients in the correlation matrix

Computation of the first-factor residuals is the second stage, which is applied by the equation

$$\rho_{ij} = r_{ij} + (a_i)(-a_j) \tag{4.3}$$

where

ρ_{ij} is the first-factor residual

r_{ij} is the correlation coefficient of two input variables i and j

a_i is the loading of the first factor in the ith variable

a_j is the loading of the first factor in the jth variable

Based on the worktable containing the residuals, the second factor loading is obtained by applying equation 4.2 to the sum of the absolute value of all the coefficients in the residual matrix. Computation of the second-factor residuals is the final procedure. This results in extracting two factors. By repeating the cycles described above, the procedure can be applied to a problem with a greater number of factors (Guilford, 1954).

Component Analysis

Contrary to common factor analysis, principle component analysis attempts to explain the total variance through the use of orthogonal (uncorrelated) principle components and makes no distinction between common and unique variance. Each factor or "component" is viewed as a weighted linear combination of the input variables, with the number of components the same as the observed variables. This resultes from the fact that the central concept of principle component factor analysis is representation and summarization. The first extracted factor typically accounts for the largest proportion of the total variance inherent in the data set. The second factor also convey quite a bit of information, though less than the first factor. Thus, each succeeding factor accounts for less and less of the total variance so that some of them explain even less variance than an input variable (Kachigan, 1991).

The first principal component can be expressed as

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{j1}X_j \tag{4.4}$$

where Y_1 represents the first extracted factor, a_{j1} is the eigenvectors comprised by the weights used to create the principal components and are scaled such that $\mathbf{a_1'a_1=I}$, X_j is the jth input variable. The variance of Y_1 is λ_1. The second principal component Y_2 is formed in such a way that its variance, λ_2, satisfies the requirements of accounting for the maximum part of the remaining variance and being orthogonal to Y_1. Components can be extracted continually until the number equals to the input variables. The eigenvalues are obtained by solving $|S - \lambda_i I| = 0$, where S is the covariance matrix. Once we have the eigenvalues, we can apply them to the equation

$$(S - \lambda_i I)a = 0 \tag{4.5}$$

to get the eigenvectors.

Criterion for Determining the Number of Extracted Factors

In this study, principal component factor analysis is employed to analyze the respondents' lifestyle characteristics because it is more convenient and can be easily applied to the FACTOR procedure in SAS system. Therefore in this section, we will focus on the determination of the optimal number of the factors to be extracted in the principal component analysis and three basic criteria will be introduced, which will also be used in the determination the lifestyle characteristic variables.

Eigenvalue

Eigenvalue corresponds to the equivalent number of variables in such a way that their variance can be explained by the derived factor. Each factor has its own eigenvalue. One frequently used rule for determining the number of factors to extract is the Kaiser-Guttman rule, which states that the number of factors to be extracted should be equal to the number of factors with an eigenvalue greater than one. The rational for choosing this particular value (one) is that a factor must have variance at least as large as that of a single original variable (Kachigan, 1991).

Scree Test

Apart from the "eigenvalue greater than one" rule, the plot of the incremental variance explained by each successive factor can also give insight into the optimal number of factors to extract. The plot is known as the scree curve. It illustrates the rate of change in the magnitude of the eigenvalues for the factors. The curve tends to decline sharply for the first few factors and then level off. The point where the curve bends is considered to indicate the maximum number of factors to extract (Kachigan, 1991).

Variance Explained

A third basis for determining the optimal factors to retain is to consider the total variance accounted for by the factors. The more factors retained, the greater the variance accounted for. However, with the purpose of data reduction, we would want to retain as few factors as possible. Therefore, the decision becomes a trade-off between the amount of parsimony and comprehensiveness we can attain. The balance that will serve our final goal becomes the basis for deciding the number of factors to extract (Kachigan, 1991).

The Factor Matrix

The end product of factor analysis is a factor matrix, with the columns representing the derived factors and the rows representing the original input variables. This matrix describes the relationship between the variables and the factors by the cell entries, which is called factor loadings. They delineate the degree to which each of the variables correlates with each of the factors with the values changing between negative one to positive one. And variables with the high loadings on a factor contribute to the meaning and interpretation of the factors (Kachigan, 1991).

Factor Score

A factor can be regarded as a variable, therefore, similar to the input variables, the factors are also measured on the same objects[6] upon which the original variables are measured. So each of our original objects also have a value on each of the derived factors. These values are called factor scores. The score represents a weighted combination of the scores of each object on each of the input variables.

Since the objective of this research is to find out whether lifestyle characteristics can be regarded as determinants of consumers' online shopping behavior, a cluster analysis is applied based on the extracted lifestyle factors. As cluster analysis is dealing with objects in the sample whereas factor analysis focuses on variables, actually, the factor scores of each respondent are submitted to the cluster analysis in order to represent the relationship between the factors and objects (Kachigan, 1991).

Rotation of Factors

When explaining factors, it is often difficult to get an impression of the underlying meaning of the factors as most of the variables have high or moderately high loadings on the first factor, and relatively few variables have high loadings on the other factors.

---

[6] Objects refer to those individual subjects in the sample whose data are obtainable.

Therefore, a redefinition of the factors, such that the loadings on the various factors tend to be either very high or very low, eliminating as many medium-sized loadings as possible, would help in the interpretation of the factors. This redefinition of the factors is accomplished by factor rotation. Many different types of rotation are available, such as varimax, quartimax, equimax, and oblimin. (Kachigan, 1991) In this study, varimax, which is the most common procedure, will be used in this study.

Cluster Analysis

Another major analytical approach employed in this research is cluster analysis. This involves a set of techniques for accomplishing the task of partitioning a set of objects into relatively homogeneous sub-groups based on the inter-object similarities. After obtaining factor scores of each respondent, the scores are submitted to the cluster analysis for the purpose of grouping the respondents based on their similarity in lifestyle characteristics. The resulting clusters will be analyzed and compared, especially through the input variables concerning online probability and demographic characteristics. By testing heterogeneity among the clusters, we can further decide whether the final clusters are significantly different from each other. On the basis of the statistically different groups, it is possible to draw conclusions as to which lifestyle characteristics may inspire customers to shop online more frequently. Furthermore, we can also compare the demographic characteristics of each cluster with the results from previous research to see whether they support each other.

There are two key assumptions with regard to cluster analysis. The most important one states that the basic measure of similarity, which is the basis of clustering, must be a valid measure of the similarity between objects. A second major assumption is that there is theoretical justification for structuring the objects into clusters (Aaker *et al.*, 1980). These two assumptions are both valid for my data set and will be used in the following analysis.

Similarity Measures

An essential step in the cluster analysis procedure is to obtain a measure of the similarity or "proximity" between each pair of objects under study. The most commonly used measure of similarity is Euclidean distance, which can be expressed as follows,

$$D_{ij}^2 = \sum_{k=1}^{p} (x_{ik} - x_{jk})^2 \tag{4.6}$$

where

D_{ij} is the distance between objects i and j

x_{ik} is the value of the kth characteristic for the ith objects

x_{jk} is the value of the kth characteristic for the jth objects

p is the number of variables

Another commonly used measure of inter-object similarity is the correlation coefficient between a pair of objects measured on several observed variables. The typical *objects* × *variables* matrix is inverted in such a way that the columns represent the objects whereas the rows represent variables. In this situation, the correlation coefficient between two columns of numbers represents the correlation, or similarity of the profile, between two objects with respect to the set of variables (Kachigan, 1991).

Non-hierarchical Clustering Approach

Non-hierarchical cluster formation is designed for disjoint clustering of very large data sets on the basis of Euclidean distances computed from one or more quantitative variables. The observations can only belong to one cluster, indicating that the mutually exclusive clusters are the end products.

Clusters are formed by a non-hierarchical clustering method through the following steps. A set of points called cluster seeds is selected as a first guess of the means of the clusters. Each observation is assigned to the nearest seed to form temporary clusters. The seeds are then replaced by the means of the temporary clusters, and the process is repeated until no further changes occur in the clusters.

When the original data set is fairly large, factor scores obtained from factor analysis can be first submitted to the non-hierarchical clustering method to get the initial seeds, which are further submitted to hierarchical clustering method to get the final clusters (SAS Institute Inc, SAS/STAT User's Guide).

Hierarchical Clustering Approach

Instead of partitioning a set of objects into a given number of mutually exclusive clusters, clusters can be formed sequentially in a hierarchical or "nested" style in which smaller clusters show up in larger ones. The output data set can be applied to TREE procedure to draw a tree diagram; as a result, we can obtain the cluster membership at any desired level.

The hierarchical clustering method in SAS includes eleven methods[7] and in this study, Ward-minimum variance method is used to get the final cluster because it gives more accurate estimation. The basic concept of Ward-minimum variance is that the distance between two clusters is the sum of squares between the two clusters added up over all the variables, and at each generation, the sum of squares within a cluster is minimized over all partitions by merging two clusters from the previous generation (SAS Institute Inc, SAS/STAT User's Guide).

[7] There are eleven clustering methods available in CLUSTER procedure, including average linkage, the centroid method, complete linkage, density linkage (including Wong's hybrid and *k*th-nearest-neighbor methods), EML (maximum likelihood for mixtures of spherical multivariate normal distributions with equal variances but possibly unequal mixing proportions), the flexible-beta method, McQuitty's similarity analysis, the median method, single linkage, two-stage density linkage, and Ward's minimum-variance method.

CHAPTER 5

DATA DESCRIPTION

Data Information and Source

The data set for this research was obtained from an online survey, which was conducted from September 27th to October 25th, 2001. This sample was from the fourth in a series of surveys intended to describe the population of Internet users who visited food and beverage sites online. First an email was sent out to these previous survey respondents inquiring whether they agree to participate another survey investigation. These respondents are mainly from the United States as well as few from other foreign countries. Two thousand two hundred and ninety of them gave a positive response as well as indicated having made an online purchase before. However, 109 provided an erroneous email address. Thus the valid sample size for this online survey decreased to 2181.

Then a questionnaire was prepared as an Internet document on the University of Maine Server and was delivered to the 2181 potential respondents by emails. The returned survey was collected by Perseus Survey Solution System and the information was transformed into an excel document. Finally, 1492 returned the questionnaire after deleting duplicated responses, indicating a response rate of 68.3%. But only 751 of them were complete and usable for factor analysis and cluster analysis.

With the intention of collecting as much information as possible, the questionnaire was composed of five parts, including information concerning to the respondents demographic characteristics, previous online behaviors, grocery shopping activities, lifestyle characteristics and the employment situation. Besides the normal questions describing respondents' demographic characteristics, the first part also included questions about their eating habits and cooking interests. An online shopping activities section mainly concerns the respondent's familiarity with the Internet, average usage of the Internet, proficiency in searching information online and previous online buying behavior. The third section asks about food purchase, e.g., respondents' regular grocery shopping activities and their satisfaction with local grocery stores. The incentives that motivate respondents' online food purchase are also investigated in this part.

The main body of the data set comes from the lifestyle characteristics section. In this part, forty-two statements are developed to measure the lifestyle dimensions drawn from other published studies in the field of consumer research. The lifestyle statements are intended to identifying the perceptions, interests and personal temperament of the respondents. A five-point Likert-type scale is used, with responses ranging from "Strongly Agree" to "Strongly Disagree". The final part lists questions related to the respondents' employment as well as its influence on the respondents' regular grocery shopping activities.

Summary of the Data

According to the 1492 valid surveys, the majority (56.3%) were aged between 40-60. Approximately 29% were forty years of age or younger, 12.1% were aged over 60. Female dominance is obvious to notice since women comprised 77.8% of the overall sample population. Over 90% of the sample has received at least some college degree. Among them, 35% have a Bachelor's degree or equivalent and 24.8% have a graduate degree. Higher than average education level is a prominent characteristic of the respondents. Concerning to the income level, the respondents are approximately equally distributed within each income category ranging from less than $20,000 to $99.999. But nearly one-fifth of them report earning more than $100,000 per year, indicating a pretty high income level of the survey respondents compared to national average. Most of the respondents have special interests in cooking, which is reflected in the cookbooks they collected. Over 60% of the sample own at least 21 cookbooks. Eating at home is another common phenomenon; 83.7% of the respondents state that they consume fewer than six meals away from home in a week. Description of the respondents' demographic characteristics is shown in Table 5.1.

Table 5.1: A Profile of Respondents' Demographic Characteristics[8]

Demographic Characteristics		Response Percentage of the Sample
Age		
	Younger than forty years old	29.7
	Forty to sixty years old	56.3
	Older than sixty years old	12.1
Gender		
	Male	19.2
	Female	77.8
Education		
	High School or less	7.6
	Some college or technical school training	30.8
	Bachelor's degree or equivalent	35.0
	Graduate degree	24.8
Income		
	Less than $20,000	3.2
	$20,000 to $29,999	4.6
	$30,000 to $39,999	8.4
	$40,000 to $49,999	8.6
	$50,000 to $59,999	9.1
	$60,000 to $69,999	5.6
	$70,000 to $79,999	6.9
	$80,000 to $89,999	5.2
	$90,000 to $99,999	5.1
	$100,000 to $150,000	12.0
	Over 150,000	6.7
Number of meals consume away from home per week		
	Less than six meals	83.7
	Seven to ten meals	11.5
	More than ten meals	2.8
Number of cookbooks currently in the household		
	Less than 20 books	37.6
	21 and more books	60.8

[8] The sum of the percentages for each variable in this table does not reach 100% because of the missing data in the 1492 returned survey. This rule is also valid for the following four tables in Chapter V.

Table 5.2 presents a profile of the respondents' online behavior as well as their attitude toward online shopping. As shown in the table, most of the respondents have a relatively long history of using the internet. Nearly ninety percent of them reported having been online for more than three years. In addition, it seems that browsing online plays an important role in the respondents' everyday life as 73.1% spend more than one hour on the Internet every day. The majority of the sample show a positive attitude toward online shopping with 88.7% reporting at least one online purchasing in the past six months and 88.8% admitting the likelihood of making another online purchase in the near future. The respondents also are interested in searching for information on the Internet as 93.1% reported they would gather information about a product or service online before making a purchase at a retail store or by mail order.

Table 5.2: A Profile of the Respondents' Online Behavior And Their Attitude Toward Shopping Online

Online Activities		Percentage
Length of the usage of the Internet		
	Less than 3 years	10.4
	3-6 years	58.2
	More than 7 years	29.6
Average hours spent online every day		
	Less than 30 minutes	4.7
	30 –60 minutes	20.8
	1-2 hours	39.3
	More than 2 hours	33.8
Previous online purchasing behavior		
	Yes	88.7
	No	9.3
Likelihood of purchasing online in the next six months		
	Likely	88.8
	Unlikely	9.3
Gathering information before making a purchase at a retail store or by mail order		
	Yes	93.1
	No	3.7

The respondents are primarily responsible for food purchases since 72% reported they did more than 80% of the household's food shopping. This phenomenon corresponds to the female dominance of the sample, indicating most women still assume the traditional role of shopping for the household's food. A profile of the respondents' regular grocery shopping is shown in Table 5.3.

Table 5.3: A Profile of the Respondents' Regular Grocery Shopping Activities

Food Shopping Activities	Response Percentage of the Sample
Time spent in the traveling from home to local grocery stores	
Less than 10 minutes	63.7
10-30 minutes	13.5
More than 30 minutes	3.6
Satisfaction with the product selection in the grocery stores	
Yes	75.3
No	22.2
Satisfaction with the product quality in the grocery stores	
Yes	80.8
No	16.9
Satisfaction with the brand selection in the grocery stores	
Yes	72.3
No	25.3

From the table, it is interesting to note that respondents are quite satisfied with their local grocery shopping. Seventy-five percent of the sample admitted their satisfaction with the product selection in the local grocery stores, 80.8% reported being satisfied with the product quality, and 72.3% stated satisfaction with the brand selection in their grocery stores. It is also convenient for the respondents to do local grocery shopping as most of them (63.7%) reported a less-than-ten-minute travel time from home to the grocery stores. Therefore, the dissatisfaction with the regular grocery shopping is not the factor that motivates the respondents to shop online. It also corresponds to the fact that online food shopping has only taken up a small percentage of the overall households' food shopping since the majority are satisfied with their local grocery shopping and still select local grocery stores as the major source of food purchasing. Other factors, such as interests in trying new things, looking for excitement, lack of availability for a particular item in the local grocery store, etc., are responsible in explaining consumer's online food shopping behavior. According to the collected data set, product availability (16.6%) and special features of the online product (11.5%) are the most often mentioned factors that attract respondents to shop on the Internet. The remaining factors are convenience (9.5%), price (5.4%), fun or curiosity (2.4%) and packaging (0.4%). The factors most important to respondent's choice to purchase online are reported in Table 5.4.

Table 5.4: Description of The Factors Most Important to Respondent's Choice to Purchase Online

Influencing Factors	Response Percentage
Convenience	9.5
Product availability	16.6
Price	5.4
Packaging	0.4
Special features of the product	11.5
Just for fun or curiosity	2.4

The lifestyle characteristics section includes statements that are commonly used to identify lifestyle attitudes and characteristics of individuals. It is the basis of this study. Table 5.5 gives a summary of responses to the lifestyle statements in the survey.

Table 5.5: General Description of Respondents' Lifestyle Characteristics

Lifestyle Statement	Agree (%)	Disagree (%)
Interests and behavior:		
1. I am often interested in theories.	73.4	24.9
2. I like to learn about art, culture, and history.	85.6	12.2
3. I would like to understand more about how the universe works.	62.7	34.9
4. I like to learn about things even if they may never be of any use to me.	83.4	14.2
5. I like outrageous people and things.	38.8	59.6
6. I often crave excitement.	45.3	52.7
7. I like a lot of excitement in my life.	39.3	58.5
8. I am always looking for a thrill.	15.5	82.0
9. I like a lot of variety in my life.	79.9	18.2
10. I follow the latest trends and fashions.	17.7	80.6
11. I dress more fashionably than most people.	16.0	82.3
12. I must admit I like to show off.	21.3	76.4
13. I want to be considered fashionable.	24.2	73.5
14. I like to dress in the latest fashion.	18.4	79.5
15. I love to make things I can use everyday.	54.3	43.9
16. I like to make things with my hands.	62.8	35.1
17. I would rather make something than buy it.	25.5	72.7
18. I am very interested in how mechanical things, such as engines, work.	40.6	57.2
19. I like the challenge of doing something I have never done before.	76.3	21.7
20. I like making things of wood, metal, or other such material.	29.4	68.5
21. I like to look through hardware or automotive stores.	42.5	55.5
22. I like trying new things.	89.3	8.9
23. I like doing things that are new and different.	73.0	24.4
24. I would like to try something even if I'm not good at it.	73.2	24.9
25. I would like to spend a year or more in a foreign country.	61.3	36.7
26. I am really interested only in a few things.	14.2	83.6
27. I must admit that my interests are somewhat narrow and limited.	9.4	88.5
28. I like my life to be pretty much the same from week to week.	19.5	78.3
29. I like to lead others.	43.8	53.9
30. I like being in charge of a group.	41.4	56.3
31. I try to avoid situations where someone else tells me what to do.	43.7	54.1
32. I would prefer not to do anything risky.	14.7	82.8
33. I avoid doing something that has the potential for rewards if it contains any risk.	8.3	89.1
34. I am often pressed for time in my day.	63.0	34.5
Beliefs:		
35. Just as Bible says, the world literally was created in six days.	17.8	78.8
36. The federal government should encourage prayers in public schools.	25.6	71.8
37. There is too much sex on television today.	45.0	52.4
38. A women's life is fulfilled only if she can provide a happy home for her family.	6.2	91.3
39. I have more ability than most people.	49.6	48.0
40. I consider myself an intellectual.	54.8	43.0
41. I think shopping online is for trend setters.	5.8	91.7
42. I think shopping online makes me adventurous consumers.	19.4	78.2

From Table 5.5, we can see that the respondents always show their special interests in learning even if the knowledge or skill will never be any use to them, which is mentioned by 83.4% of the respondents. This may correspond to the higher education level of the sample population. Interests in trying new things are also important since more than 60% of the respondents agree with the statements “I like trying new things”, “I like doing things that are new and different”, “I would like to try something even if I’m not good at it.” An online purchaser is not likable to be an extremely risk-averse person. This just coincides with the fact that 82.8% of the respondents state that they disagree with the statement “I would prefer not to do anything risky.”

It is interesting to note that the respondents take a neutral attitude toward looking for excitement and thrills as nearly half of them agree with the statement “I often crave excitement.” In addition, few respondents admit they are fashion followers, the opposite of what we have assumed. Only 17.7% of the respondents agree with “I follow the latest fashion and trend.” and 24.2% want to be considered fashionable. It is possible that online shopping is not regarded as a fashion by the respondents, which is supported by the fact that merely 5.8% of sample agree with the statement “I think shopping online is for trend setters.”

Most of the sample population report they have broad interests since more than three fourths the respondents disagree with the statements “I am really interested only in a few things,” and “I must admit that my interests are somewhat narrow and limited.” The majority of the respondents do not hold conservative opinions, which is reflected in the fact that few agree with the statement “A women’s life is fulfilled only if she can provide a happy home for her family." But most of them (63%) admit that they are always pressed for time.

Respondents also take an ambiguous attitude towards leading other people and making things with their own hands. A little bit less than 50% of the respondents agree with the statement "I like to lead others." Concerning statements related to making things by hands (statement 16, 17, 18, 19, and 20), a relatively high percentage of respondents agree with statement 16 and 19 whereas most of them disagree with the remainders. Both of these two parts need further analysis.

Employment is the final section of the survey. It tells us that 72.8% of the respondents are employed outside of the home. Most respondents (66.1%) use computer as part of their work. The majority (61.7%) drive to work every day. Although 56.3% of them indicate there is a grocery store or supermarket on the way from the their office to home, few of them (26.5%) do shopping on the way home from work.

This initial statistical analysis provides some insight as to the sample we are investigated. However, it is far from making full use of the data set. Factor analysis and cluster analysis will used in the next chapter in order to explore the relationship between the propensity of shopping on the Internet and consumer's lifestyle characteristics.

CHAPTER 6

ESTIMATION AND RESULTS

Factor analysis and cluster analysis are estimated in this chapter in order to investigate the relationship between lifestyle characteristics and the consumer's online food shopping activities. By identifying this relationship, we can identify the lifestyle characteristics associated with consumers who shop more frequently on the Internet so that online specialty food marketers can develop the corresponding advertising programs to encourage their sales. In addition, on the basis of the obtained lifestyle groups, we can compare and analyze other related characteristics, such as demographic characteristics, between each group so as to have a more general impression of their similarities and differences.

Model Estimation

A 1 to 5 scale is used in the survey, representing responses from "Strongly Agree" to "Strongly Disagree" to the lifestyle questions. But for the purpose of convenient explanation, the order of the scale is reversed so that a higher value of response indicates a stronger agreement to a particular lifestyle question.

Then the forty-two lifestyle variables were submitted to a principal component factor analysis. As discussed in Chapter 4, the basic principle of factor analysis is to express two or more variables by a single factor which captures most of the essence of the multiple items. In a sense, the multiple variables have been reduced to one factor. In principal component analysis, the extraction of the factors amounts to a variance maximizing rotation of the original variable space, the criterion for which is to maximize the variance of the factor, while minimizing the variance around the factor.

Following factor analysis, each object has its own factor scores, which are then used as an input data set to a non-hierarchical cluster analysis. The non-hierarchical cluster analysis is performed by FASTCLUS procedure in SAS and is an effective method for finding initial clusters with a standard iterative algorithm for minimizing the sum of squared distances between the cluster means. It is an efficient procedure for disjoint clustering of large data sets (SAS Institute Inc, SAS/STAT User's Guide). The obtained initial cluster seeds were further submitted to a hierarchical cluster analysis to get the final desirable clusters. The Ward-minimum variance method was employed in the hierarchical cluster analysis because it uses an analysis of variance approach to evaluate the distances between clusters. In short, Ward's method attempts to minimize the sum of squares of any two hypothetical clusters that can be formed at each step. This method tends to generate many clusters with a small number of observations. But it is good at recovering cluster structure and yields unique and exact hierarchy (SAS Institute Inc, SAS/STAT User's Guide).

Due to the unequal cell sizes, the data were tested for equal variance. Testing for homogeneity was completed by Levene's test, which "transforms the original value of the dependent variable to a dispersion variable and performs analysis of variance on this variable" (SAS Institute Inc, SAS/STAT User's Guide). In this case, question 22 is used to describe respondent's previous online food shopping activities and becomes the dependent variable to distinguish each lifestyle group. As a result, analysis of variance (ANOVA) was applied to the dispersion variable transformed from question 22.

After that, a probit model was set up to test for whether there are statistical differences among the lifestyle groups with regard to the mean value of consumer's online food shopping activities. As a result, what is modeled is the probability of online purchase, and the cluster memberships[9] become the explanatory variables. Thus seven dummy variables are used to represent the particular cluster an individual respondent belongs to. Each of the dummy variables is the symbol of a particular cluster. Based on the *p* value for each explanatory variable, the relationship between lifestyle characteristics and the consumer's online purchasing behavior can be identified.

Estimation Results

As shown in Table 6.1, ten separate factors were retained to represent the original set of lifestyle variables according to the "eigenvalue greater than one" rule. This is also confirmed by the other two criterions for determining the number of extracted factors. As reflected in Figure 6.1, the scree plot curve tends to rush downward for the first ten factors before its leveling. This indicates that the optimal numbers of factors to extract in ten. In addition, the 10-factor solution accounts for a respectable 61% of the total variance, the most appropriate balance between the variance explained and number of factors to retain.

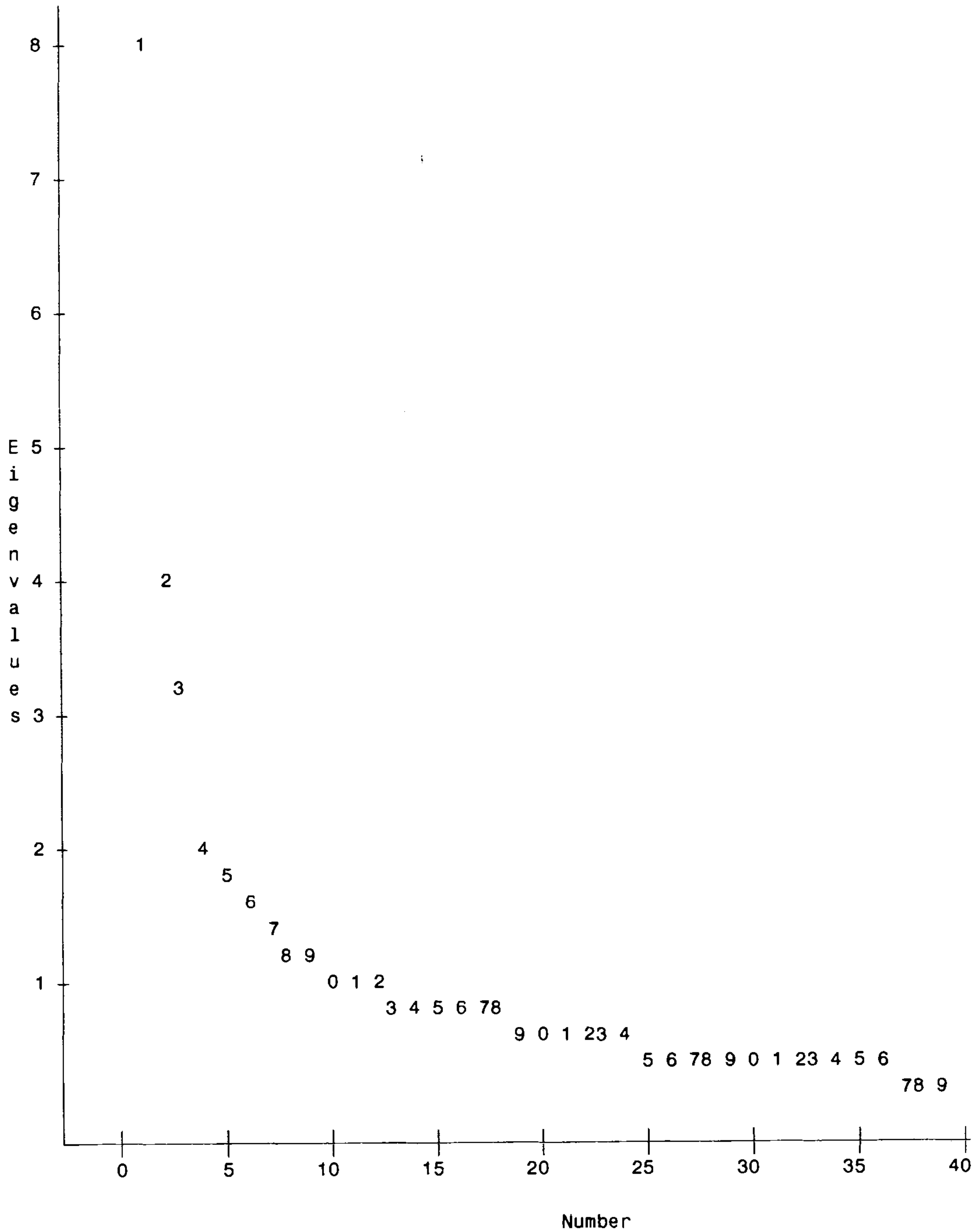
[9] Cluster membership represents to which cluster the particular respondent belongs.

Table 6.1: A Principal Component Factor Analysis of the Lifestyle Variables Collected From the 2001 Online Specialty Food Consumer Survey

Factor	Eigenvalue	Factor Loading	Lifestyle Statement
Fashion	7.92	0.78	"I follow the latest trends and fashions."
		0.81	"I dress more fashionably than most people."
		0.90	"I like to dress in the latest fashion."
		0.86	"I want to be considered fashionable."
Excitement	3.93	0.53	"I like outrageous people and things."
		0.75	"I often crave excitement."
		0.73	"I like a lot of excitement in my life."
		0.74	"I am always looking for a thrill."
New things	3.22	0.69	"I like trying new things."
		0.75	"I would like to try something new even if I'm not good at
		0.68	it."
		0.57	"I like the challenge of doing something I have never done before."
			"I like doing things that are new and different."
Self-dependent	2.07	0.81	"I love to make things I can use everyday."
		0.78	"I would rather make something than buy it."
		0.63	"I like making things of wood, metal, or other such
		0.82	materials."
			"I like to make things with my hands."
Knowledge	1.74	0.74	"I like to learn about art, culture, and history."
		0.55	"I like to learn about things even if they may never be of any use to me."
		0.57	"I would like to understand more about how the universe work."
Conservative beliefs	1.61	0.73	"Just as Bible says, the world literally was created in six
		0.80	days."
			"The federal government should encourage prayers in public
		0.62	schools."
		0.54	"There is too much sex on television today."
			"A women's life is fulfilled only if she can provide a happy home for her family."
Narrow interests	1.49	0.71	"I am interested only in a few things."
		0.73	"I must admit that my interests are somewhat narrow and limited."
Leadership	1.22	0.89	"I like being in charge of a group."
		0.89	"I like to lead others."
Mechanics	1.15	0.72	"I am very interested in how mechanical things, such as engines, work."
		0.67	"I like to look through hardware or automotive stores."
Confidence	1.08	0.63	"I have more ability than most people."
		0.54	"I consider myself an intellectual."
		0.61	"I try to avoid situations where someone else tells me what to do."

Note: The number of statements were retained based on the rule that the factor loadings should be greater than 0.50.

Figure 6.1: Scree Plot of Eigenvalues



The factor titles are based on the characteristics of the individual statements which have factor loadings greater than 0.50. Factor 1 is called fashion, and reflects the respondent's tendency to follow the latest fashions and to be considered fashionable. Factor 2 describes the respondent's interests in looking for excitement and thrills. Factor 3 is called new things. It represents the individual's preference of trying or doing things that are new to them, even if they are not good at them. Factor 4 shows the respondent's interests in making things by themselves, indicating their independence. Factor 5 is defined as knowledge and delineates the respondent's interests in learning. Conservative belief, the sixth factor, reflects traditional and conservative opinions. Narrow interests represent the respondent's limited interests and become the seventh factor. Leadership (factor 8) emphasizes the intent of the individual to lead other people and mechanics (factor 9) reflects the respondent's interests in mechanical things. The last factor, confidence, is used to depict whether the respondent is confidence in their abilities and intellects.

The factor scores of each observation, a direct output data set of factor analysis, then were submitted to FASTCLUS procedure, which resulted in 180 initial cluster seeds. Based on the initial seeds, hierarchical cluster analysis was performed and seven clusters were generated. These are shown in Table 6.2.

Table 6.2: Lifestyle Groups Obtained From Submitting the Factor Scores to the Cluster Analysis

Clusters / Factor	1	2	3	4	5	6	7
Fashion	0.55	-0.16	-0.21	-0.44	-0.74	0.09	1.03*
Excitements	0.31	-0.41	-0.21	0.04	0.63	0.52	-0.24
New things	-0.08	-0.24	-0.09	-0.45	0.81*	-0.31	0.59
Self-dependent	0.03	-0.27	0.08	0.12	-0.41	0.16	0.31
Knowledge	0.09	-0.11	0.27	0.34	0.33	-1.52*	0.04
Conservative beliefs	0.84*	0.50	0.13	-0.87*	-0.29	-0.10	-0.50
Narrow interests	0.25	0.26	-0.11	-0.32	0.19	0.07	-0.33
Leadership	-0.26	0.78*	-0.42	0.43	-0.16	-0.35	-0.66
Mechanics	-0.45	0.31	0.16	-0.15	0.13	0.57	-0.10
Confidence	0.43	0.10	-0.91*	0.09	0.39	-0.08	0.14
Sample Size	111	114	143	109	76	48	96
Percent of Sample	15.93	16.36	20.52	15.64	10.90	6.89	13.77

The names applied to the clusters were:

1. Conservative Consumers
2. Leader-oriented Consumers
3. Unconfident Consumers
4. Non-conservative Consumers
5. New Things Samplers
6. Unknowledgeable Consumers
7. New Fashion Followers

Note: The mean value with * indicates the factor that was the most significant for a particular group and therefore it can be used to represent that group.

Owing to the unequal cell sizes, tests for homogeneity were applied to the clusters. Using the general linear model framework procedure, which makes allowance for unequal cells in the ANOVA procedure, the seven clusters significantly differed from each other (p=0.0088). Levene's test was employed to test for the equal variance of the variable representing question 22, the content of which is "Have you ever purchased any food or beverage item online in the past six months?" This variable describes the respondent's online food purchase activity. Therefore, the mean value of the variable can be used to represent the likelihood of food shopping on the Internet of a certain group. The result of Levene's test indicated significant differences (p=0.0148) among the seven clusters with regard to the online food purchasing activities. These results are shown in the Table 6.3 and Table 6.4. By applying all these tests to the lifestyle groups, we can conclude that they are significantly different from each other.

Table 6.3: Using GLM Procedure to Test for Homogeneity Among the Seven Lifestyle Clusters

Source	DF	SS	Mean Square	F value	p value
Lifestyle groups	6	4.22	0.70	2.88	0.0088
Error	690	168.42	0.24		
Corrected Total	696	172.64			

Note: DF is the abbreviation of the term degree of freedom and SS is the abbreviation of sum of squares. For the purpose of reporting more accurately, the p value was retained to the fourth position after the decimal, which was the direct result of SAS program without rounding off. This criterion is also applied to Table 6.4.

Table 6.4: Levene's Test for Homogeneity of Q22Variance

Source	DF	SS	Mean Square	F value	p value
Lifestyle groups	6	0.12	0.02	2.66	0.0148
Error	690	5.16	0.01		

Note: Q22 is the variable used to reflect respondent's online food shopping activities in the past six month.

After testing the validity of the clusters, each one was given a name to emphasize the characteristics of the respondents in that particular cluster. Displayed in Table 6.2 are the respective factor loadings in each cluster. We can determine the relative importance of each factor within the lifestyle cluster. The names of the clusters were given accordingto the lifestyle factors that had the highest loading on the particular cluster. The largest of the clusters was "Unconfident Consumers" with 143 respondents accounting for 20.52% of the sample. Whereas "Unknowledgeable Consumers" and "New Things Samplers" are two minority groups, respectively representing 6.89% and 10.90% of the overall respondents. The remaining clusters do not differ significantly in size: "Conservative Consumers" (15.93%), "Leader-oriented Consumers" (16.36%), "Non-conservative Consumers" (15.64%), and "New Fashion Followers" (13.77%).

The demographic analysis has also been done to the seven lifestyle clusters, and the results are presented in Table 6.5. As can be seen from the table, the "New Things Samplers" cluster has the highest mean value on the variable representing the respondent's online food shopping behavior in the past six months, which is shown in the seventh row of Table 6.5. This implies that the respondents in this group purchase food on the Internet more frequently than those of the other groups. It also indicates that people who would like to try new things may have a higher propensity of shopping for food online.

Table 6.5: Demographic Profile of the Lifestyle Clusters

Cluster / Demo Variables	1	2	3	4	5	6	7
Gender	1.92	1.68	1.81	1.76	1.77	1.79	1.89
Education	3.60	3.93	3.77	4.17	3.94	3.48	4.02
Age	58.26	53.70	55.66	56.51	57.22	57.63	59.68
Income	6.97	8.07	8.13	8.30	7.14	7.48	8.57
Internet Usage	4.62	4.85	4.66	5.16	5.07	4.85	4.92
Time Spent Online	3.23	3.05	3.02	3.56	3.21	3.31	3.07
Online Shopping Activities	0.48	0.54	0.57	0.61*	0.71*	0.44	0.47

Note: 1. For gender variable, one represents male, whereas two represents female, therefore, the closer the mean value to one, the more males are included in the particular cluster, and vice versa.

2. For education variable, some high school or less =1; high school = 2; some college or technical school training = 3; Bachelor's degree or equivalent = 4 and graduate degree = 5, therefore, the larger the mean value of this variable, the higher the educational level of the group.

3. For income variable, the income range is from less than \$10,000 per year to more than \$100,000 per year. A larger value of the income variable indicates that the consumers in a particular cluster have a higher income level.

4. The unit of the Internet usage variable is years whereas that of time spent online is hours per day. For both variables, the higher the value of each of them, either the longer time they have been using the Internet or the more hours they spent online everyday.

5. Online shopping activities variable has two values, one represents the respondent has purchased a food item online in the past six months, whereas two indicates none of such behavior. The higher the mean value of this variable for the particular group, the more frequent that group has made a purchase on the Internet.

Therefore investigating respondents' characteristics and preferences in this cluster seem to be more attractive for Internet retailers so that they can develop the appropriate marketing programs for this special consumer groups. The marketers can advertise their products as well as provide the products' information in a way of catering to the tastes of consumers in this cluster so that they can improve their sales more efficiently. Female-dominated, middle aged, with a relatively higher education and income level than other clusters are the major demographic characteristics for this cluster. Long history of Internet usage is also a significant characteristic for this group.

Besides "New Things Samplers", "Non-conservative Consumer" is also a cluster that buys food items on the Internet more frequently than the remaining clusters. It has a relatively high value of 0.61 on the variable representing question 22, which reflects the frequency of their online food shopping activities. Therefore, the consumers in this group should become the target of online food marketers. Females also comprise the main body of this cluster with an average age of 56.5. The respondents in the "Non-conservative Consumers" cluster have an education and income level a little bit higher than the "New Things Sampler". Since online food shopping is primarily for the purpose of gift purchasing or finding something special, the price is often relatively high. The average high-income level of the respondents in this cluster indicates their ability to afford the relatively high price of online shopping. In addition, this group contains more respondents (20.52%) than any of the other cluster. All of the characteristics of the "Non-conservative Consumers" cluster imply the high potentiality of this for market segmentation by online food sellers. However, the mean values of the variable representing question 22 were almost equal for the remaining five clusters. Respondents in these clusters are infrequent online food shoppers compared to "New Things Samplers" and "Non-conservative Consumers". It also implies the potentiality of developing online specialty food markets among respondents in these groups by attracting their attention and better meeting their needs. As can be seen from Table 6.5, "New Fashion Followers" have the highest mean value on age and income, indicating respondents in this group are wealthy adults. On the contrary, "Leader-oriented Consumers" have the lowest mean value on age as well as gender showing that respondents in this group are younger and with more males. Respondents in all of the five clusters have a long history of using the Internet on average and they do spend some time online everyday.

The validity of whether "Non-conservative consumers" and "New things samplers" statistically differed from other clusters with regard to the mean value of the online food shopping activities was tested by a probit model. At the 10% significant level, the *p*-values for the coefficients of these two clusters were less than 0.1, indicating that the consumers in these two groups have more frequent online shopping behavior than those of the other groups. It is natural to expect that consumers in these two groups may have a higher likelihood to purchase something online in the future, which in turn confirmed what we have analyzed before – they are the most promising potential markets for e-tailers.

CHAPTER 7

SUMMARY AND CONCLUSIONS

Online specialty food and beverage purchasing has developed rapidly conforming to the trend of the prosperity of the Internet shopping in recent years. Researchers have been trying to figure out what factors have influenced consumers' shopping behavior in this particular field. Previous studies have focused primarily on the demographic profiles of the visitors to the food and beverage sites in order to identify the subset of this population most likely to conduct online transactions. However, these efforts have not fully explained consumers' shopping behavior on the Internet. Therefore, other attempts should be made to improve our understanding about consumers' behavior in an e-commerce environment. The focus of this study is to identify lifestyle effects on the consumer's online food shopping activities.

The data set for this research was collected from an online self-administered survey during September and October of 2001. The survey was sent out by email to 2181 respondents and 1492 were returned with a response rate of 68.3%. Based on the initial statistical analysis, we can see that the majority of the sample are well-educated, middle-aged females who are largely responsible for the household's food shopping. Lifestyle characteristics have been investigated for the first time in this area and over 90% of the respondents finished that section of the survey. According to the data, most of the respondents have interests in learning new knowledge, trying new things and doing something risky, which confirm an initial assumption.

In order to identify the relationship between the lifestyle characteristics and the online food shopping behavior, a principal component factor analysis was first used to extract ten lifestyle factors based on an "eigenvalue greater than one" criterion. These ten factors represent the forty-two lifestyle statements that are commonly used to identify the lifestyle characteristics of individuals. Then the factor scores, a direct output data set obtained from factor analysis, were submitted to a cluster analysis, using Ward's minimum-variance clustering method. Seven clusters emerged, the validity of which were confirmed by the chi-square test and GLM procedure testing for the homogeneity among groups. The results indicate that the use of lifestyle segmentation is an efficient way of analyzing the online specialty food market in a multidimensional sense.

Each cluster was given a name based on the factor that had the highest loading on that group so as to emphasize the characteristics of the particular cluster. The created lifestyle clusters are respectively, "Conservative Consumers", "Leader-oriented Consumers", "Unconfident Consumers", "Non-conservative Consumers", "New Things Samplers", "Unknowledgeable Consumers" and "New Fashion Followers". Among them, the "Unconfident Consumers" group is the largest one comprising one-fifth of the sample. It also had a second lowest value on the variable representing question 22, which reflects that the respondents of this cluster have frequent food purchasing activities on the Internet. Therefore, online food marketers should focus on this consumer niche so as to develop corresponding advertising programs to increase their sales income. "New Things Samplers" is another attractive potential market since this group of people has the most frequent online food shopping activities in the past six months. However, this is a minority group, which only accounts for approximately 10% of the overall survey respondents. This implies the limits of increasing the market shares if the online food marketers merely concentrate on this group. The respondents in the other five clusters show almost the same attitude toward purchasing food items on the Internet.

In short, lifestyle analysis provides some very important information for marketing to the online specialty food consumers. However, this study of a relatively diverse segment is only a beginning. More work is needed with the further development of online specialty food sales and more online specialty food consumers.

REFERENCES

Aaker, D. and G. S. Day: Marketing Research: Private and Public Sector Decisions, John Wiley & Sons Inc., 1980.

Alexis, Marcus, George Haines, and Leonard Simon: Consumer Information Processing: The Case of Women's Clothing, Proceedings of the American Marketing Association, ser. 28 (1968): 197-205.

Bellman, S., G. L. Gerald and E. Johnson: Predictors of Online Buying: Findings from the Wharton Virtual Test Market, Communications of the ACM. [online]. Available: http://proquest.umi.com [2001, October 22]

Bettman, J.: Information Processing Models of Consumer Behavior, Journal of Marketing Research: 370-376, 1970.

Bettman, J. and J. Jones: Formal Models of Consumer Behavior: A Conceptual Overview, The Journal of Business 45: 544-562, 1972.

Bucklin, R. and S. Gupta: Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach, Journal of Marketing Research 29: 201-215, 1992.

Burnett, J. and R. Wilkes: An Appraisal of the Senior Citizen Market Segment, Journal of Retail Banking 8, 57-64, 1985-86.

Chapman, R. and R. Staelin: Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model, Journal of Marketing Research 19: 288-301, 1982.

Donthu, N. and A. Garcia: The Internet Shopper, Journal of Advertising Research, May-June: 52-57, 1999.

Fishbein, Martin: A Behavior Theory Approach to the Relations Between Beliefs About an Object and the Attitude Toward the Object, Readings in Attitude Theory and Measurement, M. Fishbein, ed. New York: John Wiley & Sons, Inc.: 389-399, 1967.

Flath, D. and E. Leonard: A Comparison of Two Logit Models in the Analysis of Qualitative Marketing Data, Journal of Marketing Research 16: 533-538, 1979.

Food Marketing Institute (FMI), Executive Summary: The e-tail Experience: What Grocery Shoppers Think About Online Shopping, 2000. [online]. Available: http://www.fmi.org/e_business/etailexperience.html

Frank, R. E.: Brand Choice as a Probability Process, Journal of Business, 35 (January): 43-56, 1962.

Gensch, D. H.: Empirically Testing a Disaggregate Choice Model for Segments, Journal of Marketing Research 22: 462-467, 1985.

Green, P.: An AID/Logit Procedure for Analyzing Large Multi-way Contingency Tables, Journal of Marketing Research 15: 132-136, 1978.

Guilford, J. P.: Psychometric Methods, McGraw-Hill Book Company Press, 1954.

Hoffman, D. and T. P. Novak: A New Marketing Paradigm for Electronic Commerce, The Information Society 13: 43-54, 1997.

Jones, J. M. and F. S. Zufryden: An Approach for Assessing Demographic and Price Influences on Brand Purchase Behavior, Journal of Marketing, Winter: 36-46, 1982.

Jones, J. M. and F. S. Zufryden: Adding Explanatory Variables to a Consumer Purchase Behavior Model: An Exploratory Story, Journal of Marketing Research 17: 323-334, 1980.

Kachigan, S. K.: Multivariate Statistical Analysis, Radius Press, 1991.

Kamakura, W. A. and G. J. Russell: A Probability Choice Model for Marketing Segmentation and Elasticity Structure, Journal of Marketing Research (26): 379-390, 1989.

Keaveney, S. M. and M. Parthasarathy: Customer Switching Behavior in Online Services: An Exploratory Study of the Role of Selected Attitudinal, Behavioral, and Demographic Factors, Journal of the Academy of Marketing Science (29): 374-390, 2001.

Kezis, A., D. Crabtree, H. Cheng and S. Peavey: A Profile of the Specialty Food Retailing Industry in the Eastern U.S., Journal of Food Distribution Research, February: 82-91, 1997.

Kuehn, A.: An Analysis of the Dynamics of Consumer Behavior and Its Implications for Marketing Management, Ph.D. dissertation, Graduate School of Industrial Administration, Carnegie Institute of Technology, 1958.

Kuehn, A.: Consumer Brand Choice – A Learning Process? Quantitative Techniques in Marketing Analysis, Ronald E. Frank, Alfred A. Kuehn, and William F. Massy, eds., Homewood, IL: Richard D. Irwin: 390-403, 1962.

Li, H., K. Cheng and M. G. Russell: The Impact of Perceived Channel Utilities, Shopping Orientations, and Demographics on the Consumer's Online Buying Behavior. [online]. Available: http://www.ascusc.org/jcmc/vol5/issue2/hairong.html [2001, December 2]

Lumpkin, J.R. and R. Hite: Marketplace Needs of the Elderly: Determinant Attributes and Store Choice, Journal of Retailing 61: 75-105, 1985.

Malhotra, N. K.: The Use of Linear Logit Models in Marketing Research, Journal of Marketing Research, 21 (February): 20-32, 1984.

Massy, W. F., D. B. Montgomery, and D. G. Morrison: Stochastic Models of Buying Behavior, Cambridge, MA: MIT Press, 1970.

McIntosh, J.: U.S. E-Tailing Enters a New Phase, Global Online Retailing Report: 136-141, Ernst and Young.

Nicholson, W.: Microeconomics Theory: Basic Principles and Extensions, Seventh Edition, The Dryden Press, 1998.

Oates, B., L. Shufeldt and B. Vaught: A Psychographic Study of the Elderly and Retail Store Attributes, Journal of Consumer Marketing 13: 14-25, 1996.

Rosenberg, M. J., Cognitive Structure and Attitudinal Effect, Journal of Abnormal and Social Psychology, 53, 368-372, 1956.

Russ, F. A.: Consumer Evaluation of Alternative Product Models, PhD. dissertation, Carnegie-Mellon University, 1971.

SAS Institute Inc.: SAS/STAT User's Guide, Version 6, Fourth Edition, 1990.

Shufeldt, L., B. Oates and B. Vaught: Is Lifestyle an Important Factor in the Purchase of OTC Drugs by the Elderly? Journal of Consumer Marketing (15): 111-124, 1998.

Smith, M. D. and E. Brynjolfsson (2000): The Great Equalizer? Consumer Choice Behavior at Internet Shopbots, paper 137, Center for eBusiness at MIT. [online]. Available: http://ebusiness.mit.edu [2001, August 6]

Smith, M. D. and E. Brynjolfsson (2000): "Consumer Decision-making at an Internet Shopbot", paper 135, Center for eBusiness at MIT. [online]. Available: http://ebusiness.mit.edu [2001, August 24]

Sorce, Patricia, P. Tyler and L. Loomis: Lifestyles of Older Americans, Journal of Consumer Marketing (6): 13-63, 1989.

White, G.: Online Specialty Food Consumers in 1997: Who They Are and What They Buy. [online]. Available: http://www.ume.maine.edu/~specfood/demog97.html [2001, September 8]

White, G.: Catalog Specialty Food Buyers' Computer Usage and Purchasing Activity: A Case Study, Working Paper 99-03, The Retail Food Industry Center, University of Minnesota, May 1999.

White, G. and B. Manning: Characteristics of Likely Internet Specialty Food and Drink Consumers. [online]. Available: http://www.ume.maine.edu/~specfood/ama1.html 2001, November 30]

White, G.: Online Specialty Food Consumers in 2000: Who They Are and What They Buy, Draft: April 12, 2000.

White, G. and B. Manning: Convenience, Price, Product: Motivators for Online Specialty Food Consumer, Working Paper.

White, G. and B. Manning: Online Specialty Food and Beverage Consumers in 2001: Who They Are and Why They Buy Online. [online]. Available: http://www.ume.maine.edu/~specfood/papers.htm [2002, May 6]

Wildt, A. R. and J. M. McCann: A Regression Model for Market Segmentation Studies, Journal of Marketing Research 17: 335-340, 1980.

Wilkie, William L. and Edgar A. Pessemier (1973), Issues in Marketing's Use of Multi-Attribute Attitude Models, Journal of Marketing Research, 10 (November), 28-41.

Griffth, W., R. C. Hill and G. Judge: Learning and Practicing Econometrics, John Wiley & Sons Inc., 1993.

Appendix

A. Respondents' Information

1. Including yourself, how many people currently live in your household?

   1 — 1
   2 — 2-3
   3 — 4-5
   4 — 6 or more

2. Including yourself, how many of these people are 65 years old or older?

   1 — 0
   2 — 1
   3 — 2 - 3
   4 — 4 - 5
   5 — 6 or more

3. Including yourself, how many of these people are 21 years old or younger?

   1 — 0
   2 — 1
   3 — 2 - 3
   4 — 4 - 5
   5 — 6 or more

4. If you live in the USA, in which state do you live? (If you live outside of the USA, please skip to Q. 6.)

   (Click here to choose)

5. If you live outside of the USA, in which geographical region do you live? (If you live in of the USA, please skip to Q. 6.)

   (Click here to choose)

6. What is your gender?

   1 — Male
   2 — Female
   3 — I'd rather not say

7. What is the highest level of education you have completed?

- 1 Some high school or less
- 2 High school
- 3 Some college or technical school training
- 4 Bachelor's degree or equivalent
- 5 Graduate degree

8. In what year were you born? (You must be at least 18 years of age to participate in this survey)

(Click here to choose)

9. What was your total gross household income for 2000? (in U.S. $)

(Click here to choose)

10. How many meals do you usually consume away from home per week?

- 1 1-3
- 2 4-6
- 3 7-10
- 4 11-15
- 5 16 or more

11. How many cookbooks do you currently have in your home?

- 1 None
- 2 1-5
- 3 6-10
- 4 11-20
- 5 21 or more

B. Online shopping activities

12. How long have you used the Internet?

- 1 Less than 6 months
- 2 6 months - 1 year
- 3 1 - 2 years
- 4 3 - 4 years
- 5 5 - 6 years
- 6 7 years or more

13. How many hours do you spend online in an average day?

- 1 Less than 30 minutes
- 2 30 minutes - 1 hour
- 3 1 - 2 hours
- 4 3 - 4 hours
- 5 5 - 6 hours
- 6 7 hours or more

14. Have you bought anything online in the past six months?

- 1 Yes
- 2 No

15. How likely is it that you will purchase something online in the next six months?

- 1 Very likely
- 2 Somewhat likely
- 3 Uncertain
- 4 Somewhat unlikely
- 5 Very unlikely

16. Have you ever gone online to gather information about a product or service before making a purchase at a retail store or by mail order?

- 1 Yes
- 2 No
- 3 Cannot remember

C. Food purchase

17. For What percentage of your household's food shopping are you responsible?

- 1 81% -100%
- 2 61% - 80%
- 3 41% - 60%
- 4 21% - 40%
- 5 1% - 20%
- 6 0% [Go to question 22]

18. How many minutes does it take for you to travel from your home to the local grocery store where you usually shop? (one way)

- 1 Less than 5 minutes
- 2 5 - 9 minutes
- 3 10 - 19 minutes
- 4 20 - 29 minutes
- 5 30 - 59 minutes
- 6 60 minutes or more

19. How satisfied are you with the product selection in your local grocery store in which you usually shop?

- 1 Very satisfied
- 2 Satisfied
- 3 Neither satisfied nor unsatisfied
- 4 Unsatisfied
- 5 Very unsatisfied

20. How satisfied are you with the product quality in your local grocery store in which you usually shop?

- 1 Very satisfied
- 2 Satisfied
- 3 Neither satisfied nor unsatisfied
- 4 Unsatisfied
- 5 Very unsatisfied

21. How satisfied are you with the brand selection in your local grocery store in which you usually shop?

- 1 Very satisfied
- 2 Satisfied
- 3 Neither satisfied nor unsatisfied
- 4 Unsatisfied
- 5 Very unsatisfied

22. Have you purchased any food or beverage item online in the past six months?

- 1 Yes
- 2 No [Go to question 26]
- 3 Cannot remember [Go to question 26]

The following questions refer to your most recent online food or beverage purchase. Before answering, please think back to the product(s) you bought, your reasons for buying online, and the actual transaction.

23. Thinking about your most recent online food or beverage purchase, which of the following factors were important in your choice to buy online? (Please select all that apply.)

- 1 Convenience
- 2 Product availability
- 3 Price
- 4 Packaging
- 5 Special features of the product
- 6 Just for fun or curiosity

24. If you selected more than one factor, which was the most important?

- 1 Convenience
- 2 Product availability
- 3 Price
- 4 Packaging
- 5 Special features of the product
- 6 Just for fun or curiosity

25. Still thinking about your recent online food/beverage purchase, to what extent do you agree or disagree with the following statements?

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Purchasing on the Internet saved me time.	1	2	3	4	5
I could make the online purchase at a time which was convenient for me.	1	2	3	4	5
Online purchasing is easy.	1	2	3	4	5
I dislike traveling between retail stores to compare products and prices.	1	2	3	4	5
The products I purchased online are not always available in my shopping area.	1	2	3	4	5
Products similar to what I bought online are always available in my shopping area.	1	2	3	4	5
Ordering online is convenient because you can	1	2	3	4	5

change your order numerous times before completing the transaction.					
The Internet gives me access to many more business than I would have available by other means.	1	2	3	4	5
The physical location from which I made my purchase was convenient for me.	1	2	3	4	5
Online purchasing was easier than going out to do shopping.	1	2	3	4	5
The prices of the food or beverage items available at this site influenced my purchase decision very much.	1	2	3	4	5
Purchasing online saved me money.	1	2	3	4	5
I enjoy experimenting with different specialty foods and/or beverages from the Internet.	1	2	3	4	5
The Internet is convenient for gift purchases.	1	2	3	4	5
Packaging is important when the product is bought as a gift.	1	2	3	4	5
I am willing to pay more for beautiful packaging for products for my own use.	1	2	3	4	5
Food purchased online is of higher quality than food available at local retail stores.	1	2	3	4	5
Food purchased online is more safe than food purchased locally.	1	2	3	4	5
I trust that companies selling organic products online comply with accepted production standards.	1	2	3	4	5
I prefer to buy organic food.	1	2	3	4	5
I was satisfied with purchasing online.	1	2	3	4	5
I would recommend this online retailer to a friend.	1	2	3	4	5

D. Lifestyle characteristics

The following section includes statements that are commonly used to identify lifestyle attitudes and characteristics of individuals. Several of the questions asked may not seem to have any direct connection to food purchase decisions. Each of the following questions is in this survey because they have been used in other published studies of consumer research. We are interested in how responses to these statements may be related to use of the Internet for shopping. If you think any question is too invasive, please feel free to skip it, go on with the others, and submit your survey.

26. Please think about the following statements and indicate to what extent you agree or disagree with each of them.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
I am often interested in theories.	1	2	3	4	5
I like outrageous people and things.	1	2	3	4	5
I like a lot of variety in my life.	1	2	3	4	5
I love to make things I can use everyday.	1	2	3	4	5
I follow the latest trends and fashions.	1	2	3	4	5
Just as Bible says, the world literally was created in six days.	1	2	3	4	5
I like being in charge of a group.	1	2	3	4	5
I like to learn about art, culture, and history.	1	2	3	4	5
I often crave excitement.	1	2	3	4	5
I am really interested only in a few things.	1	2	3	4	5
I would rather make something than buy it.	1	2	3	4	5
I dress more fashionably than most people.	1	2	3	4	5
The federal government should encourage prayers in public schools.	1	2	3	4	5
I have more ability than most people.	1	2	3	4	5

I consider myself an intellectual.	1	2	3	4	5
I must admit that I like to show off.	1	2	3	4	5
I like trying new things.	1	2	3	4	5
I would like to try something even if I'm not good at it.	1	2	3	4	5
I am very interested in how mechanical things, such as engines, work.	1	2	3	4	5
I like to dress in the latest fashions.	1	2	3	4	5
There is too much sex on television today.	1	2	3	4	5
I would like to spend a year or more in a foreign country.	1	2	3	4	5
I like a lot of excitement in my life.	1	2	3	4	5
I must admit that my interests are somewhat narrow and limited.	1	2	3	4	5
I like making things of wood, metal, or other such material.	1	2	3	4	5
I want to be considered fashionable.	1	2	3	4	5
I think shopping online is for trend setters.	1	2	3	4	5
A woman's life is fulfilled only if she can provide a happy home for her family.	1	2	3	4	5
I like the challenge of doing something I have never done before.	1	2	3	4	5
I like to learn about things even if they may never be of any use to me.	1	2	3	4	5
I like to make things with my hands.	1	2	3	4	5
I am always looking for a thrill.	1	2	3	4	5
I like doing things that are new and different.	1	2	3	4	5
I like to look through hardware or automotive stores.	1	2	3	4	5

I would like to understand more about how the universe works.	1	2	3	4	5
I like my life to be pretty much the same from week to week.	1	2	3	4	5
I like to lead others.	1	2	3	4	5
I try to avoid situations where someone else tells me what to do.	1	2	3	4	5
I think shopping online makes me adventurous consumer.	1	2	3	4	5
I would prefer not to do anything risky.	1	2	3	4	5
I try to avoid doing something that has the potential for rewards if it contains any risk.	1	2	3	4	5
I am often pressed for time in my day.	1	2	3	4	5

E. Employment condition

27. Are you employed outside of the home?

- 1 Yes, full-time (30+ hours per week)
- 2 Yes, part-time (29 or fewer hours per week)
- 3 No (Skip to end of survey)

28. Do you use a computer as part of your work?

- 1 Yes
- 2 No

29. How often do your work hours interfere with your regular meals at home?

- 1 Always
- 2 Sometimes
- 3 Seldom
- 4 Never
- 5 Uncertain

30. How often do you drive to work?

- 1 Always
- 2 Sometimes
- 3 Seldom
- 4 Never
- 5 Uncertain

31. How many hours does it take to commute from your home to your working place?

- 1 Less than half an hour
- 2 Half an hour to one hour
- 3 1-2 hours
- 4 2 hours or more

32. Is there a grocery store or supermarket on the way from your office to your home?

- 1 Yes
- 2 No
- 3 I don't know

33. Do you usually go grocery shopping on your way home from work?

- 1 Yes
- 2 No

Thank you fro taking the time to complete this survey. Your assistance is very much appreciated. Please click "Submit Survey" when you are ready to submit your responses.

BIOGRAPHY

Xiang Xue was born in Tianjin, P.R.China on July 14, 1977. She graduated from Tianjin Yaohua Middle School in 1995. Afterward, she enrolled in the Department of International Economics and Trade, Nankai University in September 1995, and received her Bachelor of Economics degree there in July 1999. From 1999 to 2000, she has been working as an assistant manager in Shenzhen YuanTian Printing Investment Consultant Company.

In September 2000, Xiang Xue entered the University of Maine to pursue her graduate study while serving as a research assistant in the Department of Resource Economics and Policy. She is a candidate for the Master of Science degree in Resource Economics and Policy from the University of Maine in August 2002.