On Using Individual Characteristics in the MNL Latent Class Conjoint Analysis: An Empirical Comparison of the Nested Approach versus the Regression Approach

Qianqiu Zhu and Zibin Zhang

Two approaches for using individual characteristics in the MNL latent class analysis, the nested approach and the regression approach, were compared in terms of their predictive performances and model fit. When the number of individual characteristics was small, the nested approach was found to be superior, but when the number of individual characteristics was large there was no significant difference between the two approaches.

Keywords: MNL; latent class; individual characteristics; nested approach; regression approach.

Introduction

The latent class [1] MNL model is a well-established technique for representing consumer heterogeneity and segmenting the market (e.g., Andrews, Ansari & Currim 2002; Magidson & Vermunt 2003), and is especially useful in new product development and targeting research (e.g., Hauser & Urban 1977; Silk & Urban 1978; Brownstone & Train 1999), where it is used to help identify high potential products and services in innovative health services, new packaged goods, and new vehicles etc. However, a limitation of this technique as commonly used is that it uses preference data, and this type of data is often not readily available and can be difficult to obtain. So the question is whether a more easily obtained surrogate measure of preference can be used.

One possibility is to use "individual characteristics", as it has been shown that individual characteristics can be effective predictors of consumers' preferences, and are usually more readily available than conjoint preference data (Gupta & Chintagunta 1994). These "individual characteristics" usually include socioeconomic, demographic, and psychological factors. Socioeconomic and demographic factors have long been found to influence consumer choice (e.g., Horton 1979; Mittal 1994; Kalyanam & Putler 1997; Fennell et al. 2003). Furthermore, a psychological factor, "The Need for Cognition", (Cohen, Stotland & Wolfe 1955; Cacioppo & Petty 1982), has been found to influence consumers' attitudes toward a product (Haugtvedt, Petty & Cacioppo 1992), life satisfaction of college students (Coutinho & Woolery 2004), and web usage behavior (Tuten & Bosnjak 2001). Thus this study examines the use of a range of "individual characteristics" in MNL latent class analysis.

There are two ways of using individual characteristics in MNL latent class analysis. One is the nested approach (Gupta & Chintagunta 1994) and the other is the regression approach (Bucklin & Gupta 1992). The nested approach has been more popular in both academia and industry and it is already available in standard software packages (e.g., Vermunt & Magidson 2005). However, there is some evidence that, when the number of individual characteristics is large, the regression approach might be better than the nested approach (Gupta & Chintagunta 1994). However, no studies have systematically compared the two approaches, especially in terms of their performances in prediction accuracy.

[1] Also known as "finite mixture"

The purpose of the current study is to compare the two existing approaches, using individual characteristic variables in the MNL latent class analysis, in terms of prediction accuracy and model fit. The study also examines the effect of using different numbers of individual characteristic variables in the nested approach and the regression approach. Based on the findings and discussions of Gupta and Chintagunta (1994), it is expected that the nested approach will be superior to the regression approach when the number of individual characteristics is small; but for the regression approach to be superior when the number of individual characteristics is large.

The MNL latent class model

In a classical MNL latent class model without individual characteristics variables, class memberships are estimated, based on the differences in part-worth values. For a set of pre-determined numbers of classes, the probabilities that each respondent belongs to each class are calculated. The best model is selected based on the model fit criterion that was chosen by the researcher. Usually the final number of classes ranges from 2 to 6 (Green & Hensher 2003). For each respondent, their class membership and the class-specific part-worth parameters are estimated simultaneously. Using this technique, researchers can segment the market based on consumers' different part-worth values of product attributes.

The latent class model assumes that the population can be divided into Q classes or segments. The conditional probability for the nth person choosing alternative j in the ith choice situation when belonging to class q can be expressed as,

$$P(nij/q) = \exp(\beta_q' X_{ij})_i \Big/ \sum_j \exp(\beta_q' X_{ij}) \qquad (1)$$

Notice that in Equation 1, β_q is class-specific. That is, persons in different classes have different preferences, but preferences are the same within each class. The unconditional probability for the nth person choosing alternative j in the ith choice situation can be expressed as the weighted sum of the conditional probabilities:

$$P(nij) = \sum_q H_{nq} * P(nij/q) \qquad (2)$$

In Equation 2, H_{nq} is the probability of the nth person belonging to the qth class. It can be estimated along with the β vector.

Maximum likelihood method can be used to estimate the latent class model (Equation 2). The log likelihood of the unconditional probability *P(nij)* is Log (*P(nij)*). The log likelihood function for the sample (*LL*, Equation 3) is the sum of the log likelihoods over all the *I* choice sets and all the *N* consumers.

$$LL = \sum_n \sum_i \log(P(nij)) \qquad (3)$$

Latent class models have a theoretical advantage over traditional techniques of market segmentation, such as traditional cluster analysis, in that they are model-based segmentation (Magidson & Vermunt 2003). Compared to cluster analysis, the latent class models provide more accurate information on how consumers are segmented into groups. Cluster analysis segments the market based on some ad hoc definition of distance among the characteristics of consumers and product usage patterns. It is often difficult to say for certain what distance is appropriate to divide the market into segments, and the consumers near the extremes are especially difficult to deal with. The latent class models divide the consumers into groups based on their class membership probabilities that are determined simultaneously with the estimation of the part-worth values. For each consumer, the probability that she belongs to each latent class is estimated. Then, the consumer is assigned to the class that she has the largest probability. In some cases, the probability that one consumer belongs to one class is very close to the probability that he/she belongs to another class. If this should happen, the consumer will still be assigned to the class that has a greater probability. The probabilities a consumer belongs to all latent classes can also be used as the weighs of part-worth values when the researcher is interested in obtaining individual parameter estimates.

However, one problem of the latent class conjoint analysis is that it requires a conjoint experiment to be conducted and the analysis is based on the choice data. In contrast, traditional cluster analysis usually uses demographic information to segment the market. In practice, it is not always convenient to conduct a choice experiment to collect preference data. But demographic information is usually ready to use. Also, because the latent classes are unobservable, it poses a problem for practical use. The latent classes are estimated based on the part-worth values, which can only be obtained when an individual indicates his or her preferences for a product in a conjoint experiment. For those consumers whose part-worth values are unknown, it is impossible to put them into the segments and make predictions.

To address this problem, it is possible to include observable individual characteristics in the MNL latent class analysis. In this way, consumers can be segmented based on the observable individual characteristics, even if they don't indicate their preferences in a conjoint experiment. This can be achieved by relating individual characteristics to the class memberships in the latent class analysis. Once the relations of the individual characteristics and the class memberships are determined, predictions can be made based on consumers' observable characteristics.

Two Approaches for Using Individual Characteristics

The Nested Approach

The nested approach (Gupta & Chintagunta 1994) of using individual characteristics in the MNL latent class model includes two steps. The first step is to relate the class membership to a set of individual characteristics variables. It is based on the rational that a consumer's class membership is determined by their individual characteristics. Equation (4) shows a MNL structure of using individual characteristics variables to describe the class membership.

Step I: $$H_{nq} = \exp(\theta_n' z_q)^{i} \Big/ \sum_{q} \exp(\theta_n' z_q) \qquad (4)$$

H_{nq} is the probability that individual n belongs to the qth segment. θ_n is the vector of individual characteristics. Z_q is the parameter vector of the individual characteristics for the qth segment. H_{nq} can assume other functional forms that are different from the MNL in Equation 5. MNL is often adopted for its convenient form (Vermunt & Magidson 2005). The multinomial logit functional form is also used in the current study. The conditional probability for the nth person choosing the jth alternative when belonging to class q can be expressed as

$$P(j/q) = \exp(a_q + \beta_q' x_j)^i \Big/ \sum_j \exp(a_q + \beta_q' x_j) \tag{5}$$

In equation 5, the parameters are class-specific. That means, for each class, there is a different set of parameters for the part-worth values. The unconditional probability for individual n choosing alternative j can be obtained by summing up the product of Equation 4 and 5 in step two:

$$\text{Step II:} \quad P(j) = \sum_q H_{nq} * P(j/q) \tag{6}$$

This approach can be called the nested approach because the expression of the class membership using the individual characteristics is nested in the equation of the unconditional probability. The parameters for the individual characteristics will be estimated simultaneously with the class-specific part-worth values and the probabilities for class memberships.

The Regression Approach

The other approach of using individual characteristics in the MNL latent class analysis is the regression approach (e.g. Bucklin & Gupta 1992). The regression approach was proposed earlier than the nested approach, although it is less employed in previous studies. The regression approach also presents a straightforward way to use the individual characteristics in the MNL latent class analysis.

In the regression approach, the first step is to include only the product attributes into the MNL model and do the latent class analysis.

$$\text{Step I:} \quad P(j) = \sum_q H_{nq} * \frac{\exp(a_q + \beta_q' x_j)}{\sum_j \exp(a_q + \beta_q' x_j)} \tag{7}$$

The probability of the nth person choosing the jth alternative in the ith choice set is expressed as a multinomial function of product attributes. X_j is the attribute vector of the jth alternative. β_q is the vector of the part-worth parameters for the qth segment. α_q is the constant term for the qth segment. H_{nq} is the probability for the nth person belonging to the qth segment. In this step, H_{nq} is estimated solely based on the differences of the part-worth values. It is not expressed as a function of consumer characteristics. The second step is to regress the set of individual characteristics on the predicted class memberships obtained in Step I.

$$\text{Step II:} \quad \hat{H}_{nq} \sim f(z_q) \tag{8}$$

$$\hat{H}_{nq} = \exp(\theta' z_q)^{i} \Big/ \sum_{q} \exp(\theta' z_q) \qquad (9)$$

Equation 8 is a general functional form of regressing the individual characteristics variables on the predicted class memberships. Basically, the functional form *f* (.) is determined by the researcher based on prior knowledge of the relationship between individual characteristics and the class membership. Equation 9 is a multinomial logit form that will be used in the current study. It is chosen in order to be consistent with the nested approach.

The regression approach of using the individual characteristics is based on the following rationale. The factors affecting a consumer's preference for a product and their class membership may come from many sources. The observable individual characteristics are probably one of the most important sources. However, it is highly probable that other sources also affect consumer preferences. Those sources are either unobservable by the researcher or simply being overlooked in the model. It is not surprising to see that individuals with similar demographic, socioeconomic and psychometric characteristics have very different preferences toward the same product. If this should happen, it implies that other sources other than the observed ones are determining consumers' preferences. So, it may not be appropriate to express class memberships with a function of consumer characteristics alone, as in the nested approach. In the regression approach, class memberships are predicted solely based on the differences in the part-worth values. Observable consumer characteristics can be regressed on the predicted class memberships (or probabilities of class memberships) to examine their effects.

Method

A choice-based conjoint experiment was conducted via an online survey created on a professional online survey website to collect consumers' preferences for Internet services. The survey sample was 166 undergraduate students in a large Southeastern university who took same course from four different classes. Participation in the survey was voluntary and respondents who completed all the survey questions could receive extra credits in return for their participation. A total of 152 responses were usable, representing a response rate of 92%. Participants had the freedom to complete the experiment at places of their choosing within one week. The first 132 responses were used in doing the MNL latent class analysis and the last 20 responses were used to make predictions.

Survey Instrument

Sixteen choice questions with two alternatives in each question were created so that the design is orthogonal, that is, all the possible effects can be compared. Then, the sixteen questions were divided into two groups with eight questions in each group. The first eight choice questions were put at the beginning of the questionnaire in the first web page, followed by questions on individual characteristics in the second web page, and the second eight choice questions in the third web page. The purpose of splitting the choice questions was to keep up respondents' motivational levels by giving them a small number of questions at a time. For the same purpose, a progress bar was also present in the survey to show the respondents how far they had progressed with the survey. Furthermore, in order to discourage skipping questions, an announcement was made in the classes by instructors that only the respondents who completed all the survey questions could get the extra credits.

The Measurement Instrument

Five attributes of Internet services and eight variables representing individual's characteristics were used in the study. Appendix A provides an overview of these variables.

Internet Services

A literature search was conducted to identify the attributes of Internet services. Two studies on consumers' preferences for Internet service are currently available. Ida and Sato (2004) investigate consumers' preferences for broadband Internet service in Japan. The attributes in their study included cost of service, access speed, IP telephony, TV programs, service provider (NTT versus non-NTT), and Symmetry. Savage and Waldman (2004) investigate residential demand for Internet access in the US. The attributes include price, speed, always on, installation, reliability, and the feature of sharing files.

Five attributes of Internet services are used in the current study. They are price, service provider, access speed, availability of software applications, and availability of consumer support. Price and brand are commonly seen in studies using conjoint analysis (e.g. Kemperman et al., 2000) and they are probably among the most frequently used product attributes in conjoint analysis. Sometimes, price and brand are used under different names. For example, cost may be used instead of price, and service provider may be used instead of brand (Ida & Sato 2004).

Access speed is another important attribute that appears in both Internet service studies mentioned above and it will also be included in the current study under the name of speed of connection. Access speed may appear in other studies as type of connection, or physical transport to the Internet (e.g., GAO 2001). In most cases, the type of connection determines the speed of connection. For example, a dial-up connection by phone line is usually associated with a slow speed of connection and a cable modem connection is usually associated with a high speed.

Availability of software applications, such as email, anti-virus, file sharing will also be included as an attribute. File sharing has not been found to be an important attribute for US consumers, but it has the potential to become increasingly important with the widespread adoption of broad band Internet (Savage & Waldman 2004). Customer support will also be included as an attribute of Internet service in the current study. Availability of software applications and availability of customer support have also been used elsewhere for similar products, such as long distance phone services (e.g., Zubey, Wagner & Otto 2002).

"Individual Characteristics"

The "individual characteristics" included in the current study are socioeconomic, demographic, psychological, and other individual difference factors. As the sample comprises college students, the demographic and socioeconomic variables used, based on published studies, are: age, gender, race/ethnicity, and years in college (Korgen, Odell & Schumacher 2001; Jones 2002), and the primary source the respondent is using to finance his or her education (Baum & O'Malley 2003). Jones (2002) found that college students differing in age, gender, race, and years in college use the Internet differently, and college students have different places where they go online. Korgen, Odell, and Schumacher (2001) found that race/ethnicity plays a role in college students' use of the Internet. Baum and O'Malley (2003) listed various sources that college students use to finance their education and discussed how college students perceive their debt burdens. A majority of surveyed students have some forms of debt, and Baum and O'Malley (2003) implied that the source of college funding may affect college students' consumption patterns. The Need for Cognition is a personality factor, defined as "the tendency to engage and enjoy thinking" (Cacioppo & Petty 1982), and is measured with the 18-item short Need for Cognition Scale (Cacioppo, Petty & Kao 1984). Three other individual characteristics factors are also included: the number of years the respondent has used the Internet; the primary place where the respondent goes online; and the primary activity that the respondent does online (McFadden 1999; Rumbough 2001; Jones 2002).

Measures of Predictive Performance

To compare the two approaches in terms of their prediction accuracy, two measures will be used in the current study. Both of the two measures are on the model performance of predicting individual choices. The primary measure for prediction accuracy is the root mean square error between the observed P_{ij} and the predicted $\hat{P}_{ij}$ for the holdout tasks. The formula can be expressed as:

$$RMSE(P) = \sqrt{\sum_i \sum_j \frac{(\hat{P}_{ij} - P_{ij})2}{IJ}} \qquad (10)$$

A similar measure for ratings-based conjoint data is the *RMSE(Y)*, which can be found in many empirical studies (e.g. Andrews, Ansari & Currim 2002; Andrews, Ainslie, and Currim 2002). In the context of choice-based conjoint, the observed P_{ij} reflects the actual choices that consumers make. In a choice set, if 1 is used to represent an alternative being chosen and 0 not chosen, P_{ij} will assume two values: 0 and 1. The predicted $\hat{P}_{ij}$ is the predicted probability that an alternative was chosen. It is a value between 0 and 1. Then, the squared difference between the observed P_{ij} and the predicted $\hat{P}_{ij}$ is weighted by the product of the number of respondents (I) and the number of choice sets (J) each respondent has. A smaller *RMSE (P)* is desired, since the smaller the *RMSE (P)*, the better the prediction accuracy is.

The other measure for prediction accuracy is the hit rate, or the first choice correction rate (%1stCh) (Andrews, Ainslie & Currim 2002). It represents the percentage of correctly predicted choices among all the choice sets. It can be calculated with the following formula:

*%1stCh = Total number of correctly predictions / (I*J)* (11)

where %1stCh is the first choice correction rate or hit rate; I is the number of persons in the sample; J is the number of choice sets per person.

Results

Table 1 shows the results of the comparisons of the nested approach and the regression approach. MNL latent class models with the two approaches are fitted with three or eight individual characteristics included. In the three characteristics scenario, age, gender and the primary way to finance college education are used; in the eight characteristics scenario, all the individual characteristics in Appendix A are included in the latent class model. Up to seven classes were tried in order to achieve the best fit model. In all occasions, the two-class model is preferred based on the Bayesian Information Criterion (BIC) (Allenby 1990).

Overall, when three individual characteristics are used in the models, the nested approach produces higher prediction accuracy and a better model fit than the regression approach. When eight individual characteristics are included, the two approaches are not statistically different either on prediction accuracy or on model fit measures.

Table 1. Comparisons of the Nested Approach versus the Regression Approach

Measures of performance	Three individual characteristics		Eight individual characteristics	
	Nested 2 classes	Regression 2 classes	Nested 2 classes	Regression 2 classes
RMSE (P)	0.7697 [a] (0.0197)	0.8116 [a] (0.0209)	0.8031 (0.0433)	0.8089 (0.0432)
%1stCh [c]	0.5807 [a] (0.0429)	0.5329 [a] (0.0434)	0.5517 (0.0433)	0.5566 (0.0432)
BIC(LL)	1512.8955	1467.1795	1580.5695	1467.1795
LL	-690.8541	-706.7343	-697.1873	-706.7343
$\bar{\rho}$	0.6241	0.5831	0.5902	0.5831
$\bar{\rho}^2$	0.6269	0.5886	0.5939	0.5886
Number of parameters	24	11	41	11
Likelihood ratio test	31.7604 [b]		19.094	

Notes: [a] Statistically different at the 5% level.
[b] χ^2-value significant at the 5% level.
[c] Standard deviations of *%1stCh* in parentheses.

Discussion

Two measures of prediction accuracy were used in the study. They were *RMSE (P)* and *%1stCh*. When three individual characteristics were included, both measures indicated that the nested approach is statistically better than the regression approach in predicting consumers' choices using individual characteristics. Note that a smaller *RMSE (P)* value represents a better performance in prediction accuracy. However, neither of the two measures of prediction accuracy was statistically different when eight individual characteristics were included.

It is not surprising to see that the findings on *RMSE (P)* and *%1stCh* are so consistent. In obtaining the two prediction measures, individual characteristics are first used to predict consumers' class memberships. Then, based on the predicted class memberships, the probability of choosing an alternative in a given choice set can be obtained. The alternative with the highest probability is predicted to be chosen by a consumer. *RMSE (P)* is calculated based on the predicted probabilities and the actual probabilities/choices. *%1stCh* is calculated based on the predicted choices and the actual choices. So, it is not surprising that the two measures returned similar results.

A closer look at Table 1 shows that the predictive performances reflected by *%1stCh* in all cases appear to be slightly better than 50%, which is the expected hit rate when consumers make random choices. The relatively low hit rate may be due to the fact that an out-of-sample prediction was made in the current study.

An interesting finding is that the nested approach with three individual characteristics has the highest prediction accuracy. This implies that simply increasing the number of individual characteristics will not increase the prediction accuracy of an MNL latent class model.

Model fit was measured by the log likelihood (*LL*) of the model. Because two-class models were chosen for both the nested and the regression approaches, a likelihood ratio test can be conducted to compare the model fit of the two approaches (e.g., Gupta & Chintagunta 1994). When three individual characteristics are included, the model fit is statistically better for the nested approach than for the regression approach. The improvement, measured by $\bar{\rho}^2$, is 0.0383, which is relatively modest. However, the difference between the two approaches disappears when eight individual characteristics are included.

In summary, the performances of the nested approach and the regression approach appear to depend on the number of individual characteristic variables. When the number of individual characteristics is small, the nested approach is better; when the number of individual characteristics is large, the superiority of the nested approach disappears. In this study, the regression approach was actually better than the nested approach when eight individual characteristics are included, although the difference was not statistically significant.

References

Allenby, GM. (1990). Hypothesis Testing with Scanner Data: The Advantage of Bayesian Methods, *Journal of Marketing Research*, 27, 379-389.

Andrews RL, Ainslie A & Currim IS (2002). An Empirical Comparison of Logit Choice Models with Discrete Versus Continuous Representations of Heterogeneity. *Journal of Marketing Research*, 39 (November), 479-487.

Andrews RL, Ansari A & Currim IS (2002). Hierarchical Bayes vs. Finite Mixture Conjoint Analysis Models: A Comparison of Fit, Prediction, and Partworth Recovery, *Journal of Marketing Research*, 39 (February), 87-98.

Baum S & O'Malley M (2003). *College on Credit: How Borrowers Perceive their Education Debt (Results of 2002 National Student Loan Survey)*. Braintree, MA: Nellie Mae.

Brownstone D & Train K (1999). Forecasting New Product Penetration with Flexible Substitution Patterns, *Journal of Econometrics*, 89(1-2), 109-129.

Bucklin RE & Gupta S (1992). Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach, *Journal of Marketing Research*, (May), 201-215.

Cacioppo JT & Petty RE (1982). The Need for Cognition, *Journal of Personality and Social Psychology*, 42, 116–131.

Cacioppo JT, Petty RE & Kao CE (1984). The Efficient Assessment of Need for Cognition, *Journal of Personality Assessment*, 48, 306-307.

Cohen AD, Stotland E & Wolfe DM (1955). An Experimental Investigation of Need for Cognition. *Journal of Abnormal and Social Psychology,* 51, 291–294.

Coutinho SA & Woolery LM (2004). The Need for Cognition and Life Satisfaction among College Students, *College Student Journal*, 38(2), 203-206.

Fennell G, Allenby GM, Yang S & Edwards Y (2003). The Effectiveness of Demographic and Psychographic Variables for Explaining Brand and Product Use, *Quantitative Marketing and Economics*, 1(2), 223-244.

GAO (2001). *Characteristics and Choices of Internet Users.* Washington (DC): United States Department of Commerce, http://www.gao.gov/new.items/d01345.pdf. Accessed 25 February, 2005.

Greene WH & Hensher DA (2003). A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit, *Transportation Research: Part B*, 37(8), 681-699.

Gupta S & Chintagunta PK (1994). On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models, *Journal of Marketing Research*, 31 (1), 128-136.

Haugtvedt CP, Petty RE & Cacioppo JT (1992). Need for Cognition and Advertising Understanding the Role of Personality Variables in Consumer Behavior, *Journal of Consumer Psychology*, 1, 239-260.

Hauser JR & Urban GL (1977). A normative methodology for modeling consumer response to innovation, *Operations Research*, 25(4), 579-619.

Horton RL (1979). Some Relationships between Personality and Consumer Decision Making, *Journal of Marketing Research*, 16(2), 233–246.

Ida T & Sato M (2004). *Conjoint analysis of consumer preferences for broadband services in Japan*. Manuscript submitted for publication.

Jones S (2002). *The Internet Goes to College: How Students Are Living in the Future with Today's Technology*. Pew Internet & American Life Project, http://www.pewinternet.org/reports/toc.asp?Report=71. Accessed 25 February, 2005.

Kalyanam K & Putler DS (1997). Incorporating Demographic Variables in Brand Choice Models: An Indivisible Alternatives Framework, *Marketing Science* 16, 166–181.

Kemperman AD, Aloys AM, Borgers, WJ, Oppewal H, & Timmermans HJP (2000). Consumer Choice of Theme Parks: A Conjoint Choice Model of Seasonality Effects and Variety Seeking Behavior, *Leisure Sciences*, 22(1), 1-18.

Korgen K, Odell P & Schumacher P (2001). Internet Use Among College Students: Are There Differences By Race/Ethnicity?, *Electronic Journal of Sociology*, 2(2), http://www.sociology.org/content/vol005.003/korgen.html. Accessed 25 February, 2005.

Magidson J & Vermunt JK (2003). A Nontechnical Introduction to Latent Class Models, *DMA Research Council Journal*, http://www.statisticalinnovations.com/articles/lcmodels2.pdf. Accessed 25 February, 2005.

McFadden AC (1999). College Students' Use of the Internet, *Education Policy Analysis Archives*, 7(6), http://epaa.asu.edu/epaa/v7n6.html. Accessed 25 February, 2005.

Mittal B (1994). An Integrated Framework for Relating Diverse Consumer Characteristics to Supermarket Coupon Redemption," *Journal of Marketing Research*, 31, 533–544.

Rumbough T (2001). Paper Mills, Pornography, and Pirating: College Students and the Internet: A survey at Bloomsburg University of Pennsylvania Reveals that Students Use the Internet for Sometimes Unethical Purposes, *Educause Quarterly*, 4, 59-60.

Savage S & Waldman DM (2004). United States Demand for Internet Access," *Review of Network Economics*, 3(3), 228-247, http://ssrn.com/abstract=603162. Accessed 25 February, 2005.

Silk AJ & Urban GL (1978). Pre-test-market evaluation of new packaged goods: a model and measurement methodology, *Journal Marketing Research,* 15 (May), 171-191.

Tuten TL & Bosnjak M (2001). Understanding Differences in Web Usage: The Role of Need for Cognition and the Five Factor Model of Personality, *Social Behavior & Personality,* 29(4), 391-399.

Vermunt, JK & Magidson J (2005). *Technical Guide to Latent GOLD Choice 4.0: Basic and Advanced.* Belmont, Massachusetts: Statistical Innovations Inc.

Zubey ML, Wagner W & Otto JR (2002). A Conjoint Analysis of Voice over IP Attributes, *Internet Research,* 1(2), 7-15.

Qianqiu Zhu is a Senior Statistical Analyst at Experian, Atlanta, GA, USA, and Zibin Zhang is an Assistant Professor in the College of Economics, and China Academy of West Region Development, both at Zhejiang University, Hangzhou, China.

Appendix A. Description of Variables

	Name	Description	Measurement
Dependent variable	PREF	Consumer preference for Internet service, measured by the probability of one profile being chosen	Being chosen: 1 Not being chosen: 0
Independent variables			
	PRICE	Price of Internet service	\$25.99: 0 \$35.99: 1
	CONNECTION	Speed of connection to the Internet	Dial-up: 0 High-speed: 1
	COMPANY	Whether the service provider is a national company or a local company	Local: 0 National: 1
Attributes of Internet service	APPLICATION	Whether there are software applications such as anti-virus, firewall, email, file sharing, etc.	No: 0 Yes: 1
	SUPPORT	Availability of customer support	Limited to No: 0 24/7: 1
Classification variables			
	AGE	Age of the respondent	A numerical value reported by the respondent
Socioeconomic and demographic characteristics of consumers	GENDER	Gender	Male: 0 Female: 1
	HOURS	Number of credit hours earned so far	Less than 30 hours: 0 30 to 60 hours: 1 60 to 90 hours: 2 90 hours and above: 3
	RACE	Ethnicity of the respondent	African Americans: 0 Asian: 1 Hispanic: 2 White: 3 Other: 4
	FINANCE	The primary source the respondent is using to finance her college education	Family (parents, etc.): 0 Hope scholarship: 1 Other scholarships and or grants: 2 Loans: 3 Other: 4
	EXPERIENCE	Number of years the respondent has used the Internet	Less than 1 year: 0 1 to 3 years: 1 3 to 5 years: 2 5 years or more: 3
	PLACE	The primary place where the respondent goes online	Apartment/Dorm: 0 On campus (Computer labs, libraries, etc.): 1 Free public facility off campus: 2 Off campus facility that charges: 3 Other: 4
	ACTIVITY	The primary activity that the respondent does online	Communication (email, instant messaging,chat rooms, etc.): 0 Learning (homework, research, online courses, etc.): 1 Entertainment (play games, hobbies, etc.): 2 Information (search for info, read news, etc.): 3