A Comparative Study of Data Gathering Procedures in Conjoint Measurement

著者	Ogawa Kohsuke, Hattori Shota, Fukushima Tsunehiro
出版者	法政大学産業情報センター
journal or publication title	グノーシス : 法政大学産業情報センター紀要
volume	2
page range	18-33
year	1993-03-31
URL	http://hdl.handle.net/10114/00020855

A Comparative Study of Data Gathering Procedures in Conjoint Measurement (1)

Kohsuke Ogawa, Shota Hattori, Tsunehiro Fukushima.

ABSTRACT

An experiment is designed for testing validity and reliability of two data gathering procedures in conjoint measurement. Computer-interactive Adaptive Conjoint Analysis (ACA in short) and the conventional Full-Profile method (FP in short) are among those compared for predictive performance.

After responding to four questionnaires, two data collection procedures each for two product categories (chocolate and soft drink), in a computer-assisted session, two hundred and six respondents picked up their most favorite brand(s) from a set of brands with relatively high shares in the market.

For soft drink category, partworths of product attributes are estimated for price, manufacturer, brand category, and size of container. For chocolate, importance weights are estimated for price, maker, taste, and product form.

Average correlation coefficients between parameter estimates derived from the different data collection procedures (ACA and FP) are quite high; 0.52 on the average for both product categories, mostly above 0.65 for individual respondents.

Using parameter estimates, total utility scores could be calculated for the brands presented at the final stage of computer interview. Then, the first choice could be predicted and matched with the brand actually picked up by each respondent. "Batting Average" for FP method is 53.9%, which is fairly higher than 44.7% for ACA procedure in predicting the choice of chocolate. However, ACA with an average of 45.7% could hit the right cans of Cola, Tea, or Orange Juice better than FP only with an average of 40.4%.

We recommend that researchers would better make use of ACA against FP, when there are many attributes and/or profiles, since interviewing with ACA is much easier than that with FP.

1. Introduction

Conjoint Analysis, sometimes called Conjoint Measurement, is a set of well-known techniques to evaluate a product concept in an early development stage of new product design, to select some salient attributes and/or determine its optimal price level in a later phase of product development. Since its methodological inception by Kruscal (1965), a lot of works have been done so far for performance comparison among alternative data gathering procedures, parameter estimation techniques, and methodological extensions among others. Green and Srinivasan (1978, 1990) summarized the state of art in conjoint study.

Wittink and Cattin (1989) reported that conjoint analysis has been applied to more than 1,000 cases in private business and public institutions. Case numbers in commercial applications does not seem to decrease even now. Popularity of conjoint analysis in our marketing research society is partly attributable to its easy-to-use feature as a research technique, whatever alternative data gathering or statistical procedures you take, compared to other techniques in marketing research. Recent technological development in personal computing will facilitate its further diffusion of the conjoint method in commercial applications.

In a typical conjoint experiment, a set of product profiles, often with pictures or description of assumed product features, are presented for respondents. They are often requested to rank-order the entire profiles, i.e., the most to the least preferred in Full-Profile Method. With such a ranking data collected, then researchers can calibrate attribute importance vector, called partworths, using any one of statistical procedures for partworths estimation, e.g., OLS, MONANOVA, LINMAP, or LOGIT.(2)

Rapid technological change in the personal computer industry enables researchers to depend on a PC-based interactive interviewing system in order to obtain partworths values. A couple of software packages are now available for a field research at a minimal cost.(3) PC software basically uses a logic of modified version of paired comparison method called Trade-off Method originally developed by Johnson (1975).

In a computer-assisted session, respondents are instructed to successively answer such questions as to "which profile is the more important and to what extent," while a pair of profile descriptions appearing on a computer screen. A stopping rule is set in an algorithm which makes a respondent feel difficulty in choosing the better alternative on the screen after some rounds. At that time, the computer automatically stops generating the next pair of competitive profiles, then it calibrates partworths estimates successfully.

The trade-off method has an apparent advantage over the full-profile method for respondents, i.e., easy to respond and convenient for re-analysis and a market simulation. However, as far as computer software is concerned, its convergence property is not well-known, and some practitioners care about the reliability and validity. They wonder if the subsequent market simulation might not lead to the right solution for the right marketing decision.

Hence, we planned a conjoint experiment by which we can check out superiority of alternative data gathering procedures.(4) We compare the trade-off method and full-profile method in terms of reliability and external validity. First, a comparison is made to see similarity of partworths calibrated by different data-gathering procedures. Next, each procedure is evaluated in view of how it can predict an actual consumer choice, correctly.

For a descriptive convenience, we denote the full-profile method by FP, the trade-off method by ACA, because of the name of software package, Adaptive Conjoint Analysis, which we used in the experiment.

2. Outline of Conjoint Experiment

Product Category and Attribute

After enumerating then screening several candidates for product categories, chocolate and soft drink were chosen for our experiment. We assumed that young respondents must be familiar with these product categories, especially with soft drink. Cost at a free sample is another reason for this decision.

Gender effect in choosing a chocolate was expected. Women like to buy a chocolate more frequently than men. Therefore, women usually have a larger evoked set of chocolate brands. The forty-one percent of male participants answered they never bought a chocolate in a month.

For soft drink category, we constructed a product profile with four attributes, i.e., price (three levels), manufacturer (five makers), brand category (five sorts), and the size of container (three levels). For chocolate, a profile description is made of attributes, i.e., price (three levels), manufacturer (four makers), taste (four levels), and the product form (two kinds). Regarding the FP method, twenty-five profile cards for soft drink were generated by Addelman's orthogonal factorial design, and sixteen cards for chocolate.(5) The attribute levels are exhibited in Table 1.

Table 1
List of Attributes and Levels

A. Soft drink Category

Attribute	Price	Maker	Category	Size
Level 1	¥80	Coca Cola	Carbonated	125 ml
Level 2	¥100	Ohtsuka	Isotonic	250 ml
Level 3	¥120	Suntory	Black Tea	350 ml
Level 4		Kirin	Coffee	
Level 5		UCC	Orange Juice	

B. Chocolate Category

Attribute	Price	Maker	Taste	Form
Level 1	¥100	Lotte	Milk	Plate
Level 2	¥150	Meiji	White	Flake
Level 3	¥200	Morinaga	Bitter	
Level 4		m & m	Nuts	

Research Plan

The experimental research was executed during two weeks of early June in 1991 at the Tokyo branch and Kumamoto headquarters offices of Kozokeikaku Engineering, Inc. One hundred and fifteen undergraduate students in the first author's class of "marketing science" and one hundred and six young employees in the second author's company sat down to a Macintosh in the booths to answer a sequence of specially programmed questionnaire. Out of 221 samples, only two hundred and six (133 men and 73 women) were effectively submitted to the estimation stage.

In 1990 a year before, we developed and tested a NEC PC98 version of Japanese ACA by Kozokeikaku Engineering, Inc. At that time, a set of profile cards made of craft papers was requested to manually arrange in order of individual preference, while a respondent hit the key board in answering an ACA questionnaire.(6) In the experiment of the year 1991, however, two procedures (ACA and FP) were packaged into a series of PC questionnaires. A participant was asked to respond to the questions on the screen of PC with operating a mouse for both ACA and FP sections.

Since we adopted two product categories at a time for testing and intended to obtain two different data set for the same subjects, the resulting number of patterns in a sequential presentation is four types by combination. We split the whole sample into four groups in order to prevent "order bias" in presentation, i.e., to neutralize the order effect in presenting products and collecting data procedure upon experiment.

Table 2

Four Patterns of Response Order

Pattern 1:	FP(Chocolate) → FP(Soft drink) → ACA(Chocolate) → ACA(Soft drink)
Pattern 2:	FP(Soft drink) → FP(Chocolate) → ACA(Soft drink) → ACA(Chocolate)
Pattern 3:	ACA(Chocolate) → ACA(Soft drink) → FP(Chocolate) → FP(Soft drink)
Pattern 4:	ACA(Soft drink) → ACA(Chocolate) → FP(Soft drink) → FP(Chocolate)

Respondent were evenly assigned to any one of patterns above. After finishing additional behavioral and demographic questions, each respondent was requested to pick up his/her most favorite brand from the real products on the table. Four brands for chocolate and five brands for soft drink were offered to eat and drink for free when finishing a PC interview.

Explanatory Factors and Scaled Measurements

Factors affecting brand preference are mentioned in the first column of Table 3 below. Among them are respondent's sex, product category, knowledge about product class, consumption experience, and the degree of variety-seeking behavior.

Taking a strong advantage of automatic data recording characteristics by personal computer, we counted a total as well as partial response time in seconds for each respondent, which in turn represents a measure of ability to handle a PC, information overload, or attitude toward the experiment. We may call these measures intermediate variables, since these parameters, relying on the demographic and behavioral factors, might have an effect on parameter's reliability and predictive validity.

In contrast, we calculated some scaled measures based on the choice data by respondents. To see the difference in partworths estimates and the actual choice of brands, we calibrated three performance measures as follows:

(1) <u>Consistency Measure</u>:

Simple correlation coefficient between partworths calibrated by ACA and FP,

(2) <u>Internal Reliability Measure</u>:

Reliability Measure for ACA; likelihood ratio index ρ for FP,

(3) <u>Predictive Validity Measure</u>:

Batting average in prediction for the actual choice in comparison with theoretical (first) choice based on partworths estimates.

<u>Table 3</u>
<u>Explanatory Variables and Performance Measures</u>

Exploratory Variables	Data Collected Intermediate Parameters	Performance Measures
	Preference ranking data	FP (MONANOVA & LOGIT) partworths estimate
Knowledge about product feature	Paired comparison data	ACA
	Information overload	(1) Consistency
Consumer characteristics: age, sex, variety-seeking, expertise in PC, buying habit	Response time	(2) Reliability
	Attitude measure	(3) Predictability
	A brand chosen from alternative products	

Manipulation Factors: Product, procedures, order in presentation

3. Results and Findings

Response Time

Respondent's overall attitudes toward the computer interview were quite nice. Ninety-two percent of participants told that they were so impressed by "a computer game" that they enjoyed the computer interview. Mean total response time was about forty minutes. Most respondents, one hundred and seventy-three out of 206 effective samples took time within a range of thirty-five to fifty-five minutes, with the shortest taking fifteen minutes and the longest eighty minutes.

We split the total response time into four parts in a row, product category by data-gathering procedure. As was expected, the section of soft drink by ACA took the longest time of twelve and half minutes on the average, the shortest of seven minutes for chocolate by ACA; nine minutes for chocolate by FP, eight minutes for soft drink by ACA. As far as response time is concerned, ACA is the better research option for obtaining preference data in conjoint measurement by about ten percent margins.

Partworths Estimates

A snake plot of mean partworths for soft drink by FP is shown in the Panel A of Figure 1. Very similar pattern of the snake plot for ACA can be seen in the Panel B just below.(7) Both estimates are quite similar not only in magnitude but also in pattern. A subtle difference, if any, might be found in price sensitivity. The slope of the line connecting ¥80 through ¥100 with ¥120 is much steeper for ACA than for FP. It seems to us that ACA tends to give a more price sensitive solution than FP, although we have to accumulate more cases in a variety of products and consumer segments.

The reason is simple. It can be explained by a principle of consumer's information-processing. Faced with a lot of profiles like in a case of soft drink for FP, a respondent appears to take such an information-processing strategy that, at first she focuses on her most salient attribute, which becomes a pivot for sorting out her cards in order of preference ranking. Price is just one of the candidates in that case. However, in a situation of paired comparison for ACA, price plays quite a different role in each stage of pairwise judgment. With only two profiles for comparison at a step, she may always pay attention to price attribute.

Importance weights of attributes for soft drink seems to be allocated evenly. One the other hand, taste is the most important attribute in choosing a chocolate, next comes price factor in selecting a chocolate. The other attributes, maker and product form, might not be so critical in a choice of chocolate. Figure 2 with the Panels A and B describes the observation above. Again the basic pattern of

Figure 1 (A)
Snake Plot of Mean Partworths
for Soft drink by Full-Profile Method

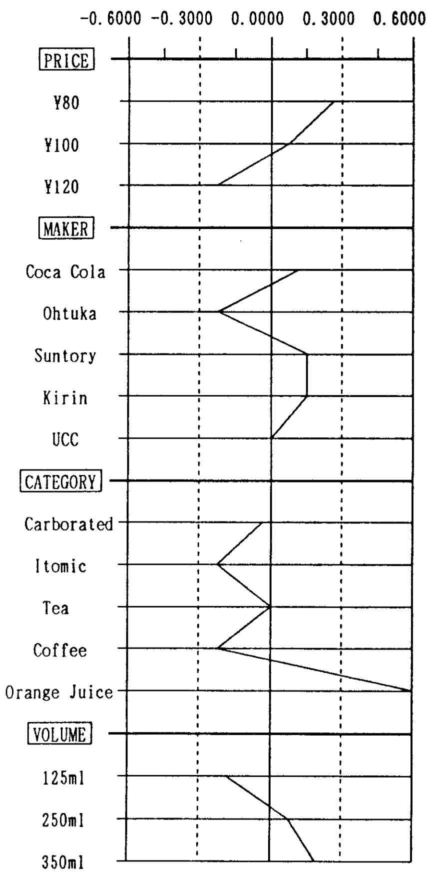


Figure 1 (B)
Snake Plot of Mean Partworths
for Soft drink by Adaptive Conjoint Analysis

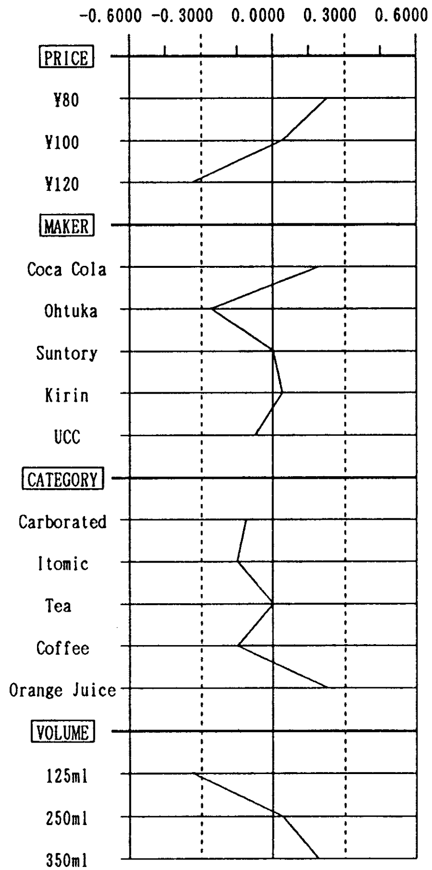


Figure 2 (A)
Snake Plot of Mean Partworths
for Chocolate by Full-Profile Method

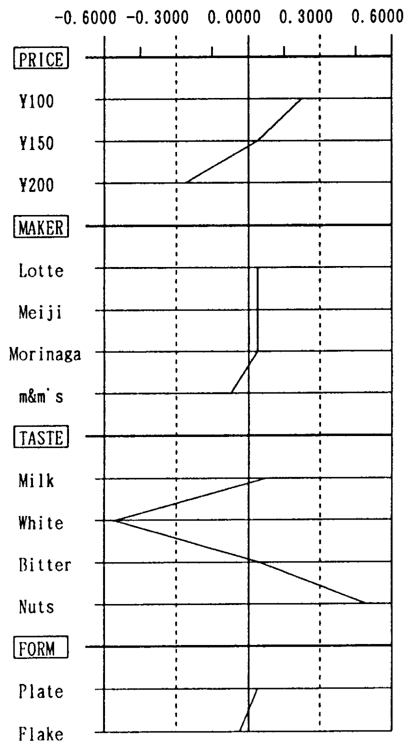
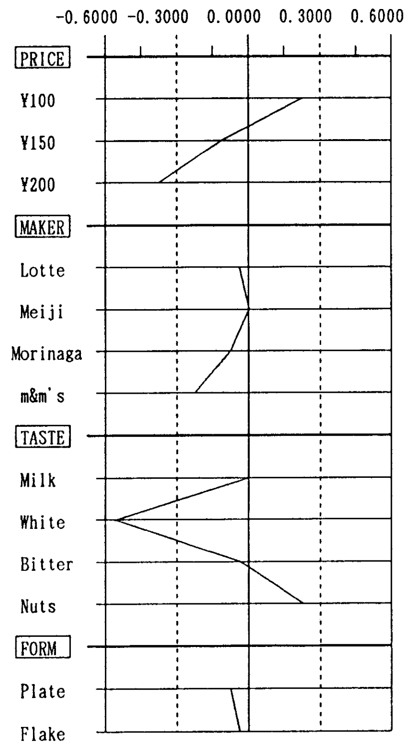


Figure 2 (B)
Snake Plot of Mean Partworths
for Chocolate by Adaptive Conjoint Analysis



importance weights looks alike.

Correlation between Partworths by ACA and FP

Simple correlation between partworths by FP and ACA was calculated on an individual base for chocolate and soft drink.(8) A frequency distribution of individual correlation coefficients between partworths independently estimated by FP and ACA can be seen in the Panels A and B of Figures 3. Average correlations of partworths over whole sample, which we call consistency measure later in this article, is 0.52 incidentally same for both chocolate and soft drink.

We might say that most respondents are moderately consistent in a computer interview. A few respondents, approximately ten percent, appear to forget their answers in the subsequent section of the interview, since it often takes about ten minutes between the sections of ACA and FP. Our recommendation is that, when respondent's partworths correlations are negative, we had better throw out these their data for a market simulation.

Other than FP and ACA, we could estimate the partworths by Kruscal's MONANOVA, which was coded by CVA in our convention. Hence in total we had three set of parameters for a product class. Triangular vis-a-vis comparisons are possible for calculation. Mean correlation coefficients are presented in Table 4.

Judging from a range of 0.52 to 0.95, partworths are highly correlated one another for all three matched pairs. In sum, correlation for chocolate is better than for soft drink category because of large degrees of freedom.

Table 4
Simple Correlation of Individual Partworths
Calibrated by Comparative Data-Gathering Procedures

Data-Gathering Procedure	Categry	Average Correlation by Respondents
A C A v s C V A	Chocolate	0.53
	Soft-drink	0.55
C V A v s F P	Chocolate	0.87
	Soft-drink	0.95
A C A v s F P	Chocolate	0.52
	Soft-drink	0.52

Figure 3 (A)
Frequence Distribution of Correlation Coefficients between Partworths Vectors (Soft drink)

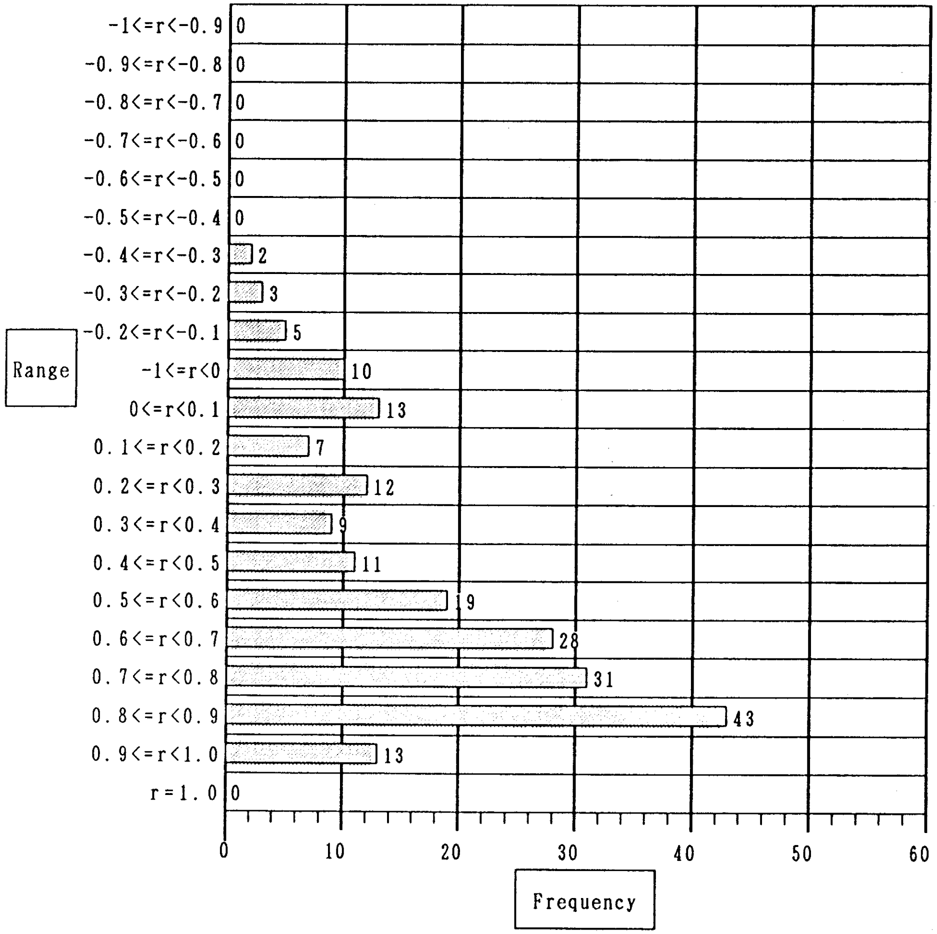
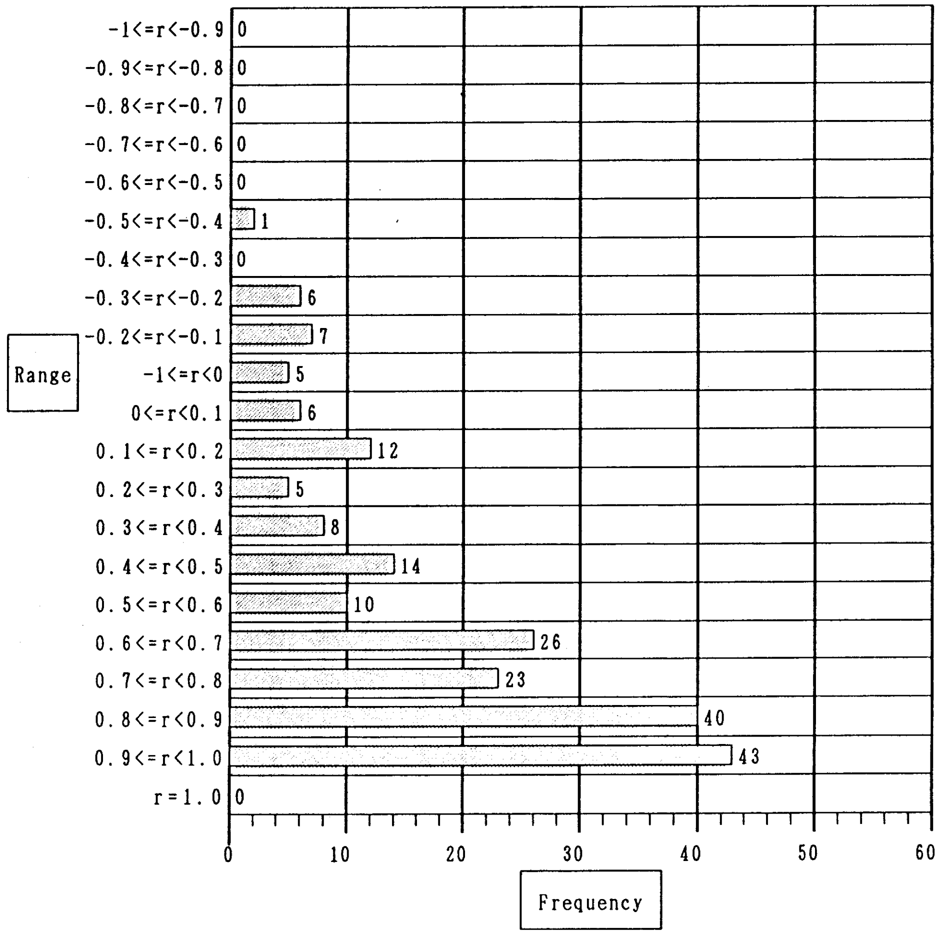


Figure 3 (B)
Frequence Distribution of Correlation Coefficients between Partworths Vectors (Chocolate)



Likelihood Ratio Index and Reliability Measure

For Full-Profile method, likelihood ratio index ρ is used for internal validity check. With higher roh coefficient, for example, at the level of 0.8 to 1.0, we could almost recover the original rank ordering of product profiles resulting from estimated partworths. With lower roh coefficient, for instance, less than 0.3, we would obtain quite a different rank order from the reported one by respondents.

For Trade-off method, on the other hand, Reliability Measure is automatically calibrated in the last part of ACA section.(9) This measure is a sort of coefficient of determinant in OLS, where the stated preference is regressed on the theoretical partworths calibrated based on a logic of ACA. However, as far as our experience in two categories for a small sample is concerned, reliability measure might not be a good fitting index for validity check. We could not find any statistically significant relationship between ACA's reliability measure and our consistency measures. Any positive correlation could not be observed in our experiment for both product categories.

In contrast, likelihood ratio index has a positive correlation with consistency measure. We might conclude that ratio index reflects reliability of reported data, therefore stability of partworths estimates within an individual respondent. A scatter plot of roh over consistency measurement is give in Figures 4 and 5. Simple correlation is about 0.41 for chocolate and about 0.45 for Soft drink.

Hit Rate for Actual Choice by Conjoint Model

At the final stage of experiment, a participant was solicited to choose his most favorite brand from a set of real products displayed on the table before PC. Non of them rejected our offer. At first, he picked up a bottle of soft drink from five brands with relatively high shares in the market. Then he took out a package of chocolate from four brands after responding a short question regarding demographics in a few minutes.

A choice set was carefully constructed, since we would like to balance the mixture of attribute levels. That is, any attribute level must appear at least once in any one of products offered, exclusive of price levels which we could not control in that experimental setting. The followings in the Table 5 are the choice set compositions of actually presented brands to participants.

In the leftest columns of Table 5, we can see the number of first choice respondents for each brand. At that time, Kirin just launched a new brand, Kirin's Tea ("Gogo-NO-Kocha" in Japanese), which was included in the choice set. Thus, there might be a campaign effect as well as a hollow effect of Kirin's Tea for such a high share taken. For chocolate category, any brand did not bite a big portion of the market pie. It can be said that chocolate, in a generic sense, tends to be

Figure 4

Scatter Plot of Sqrt(roh) v.s. Correlation

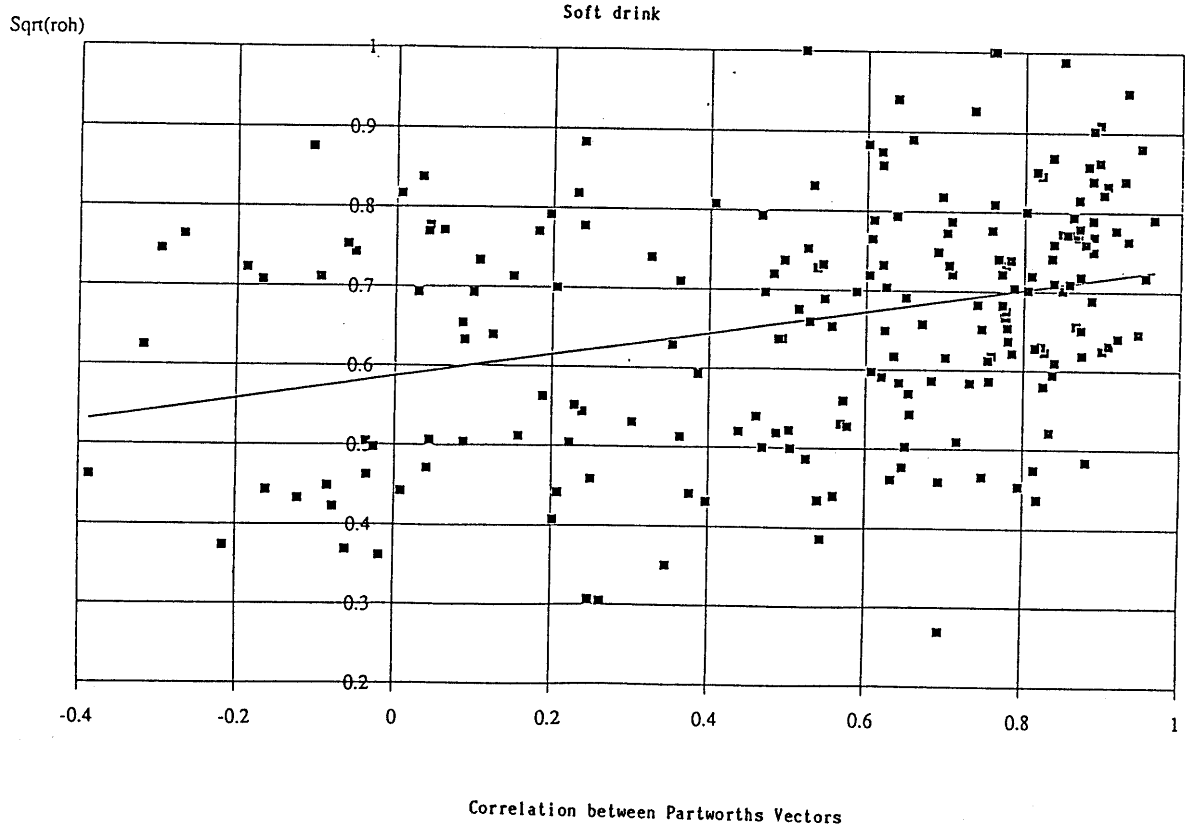
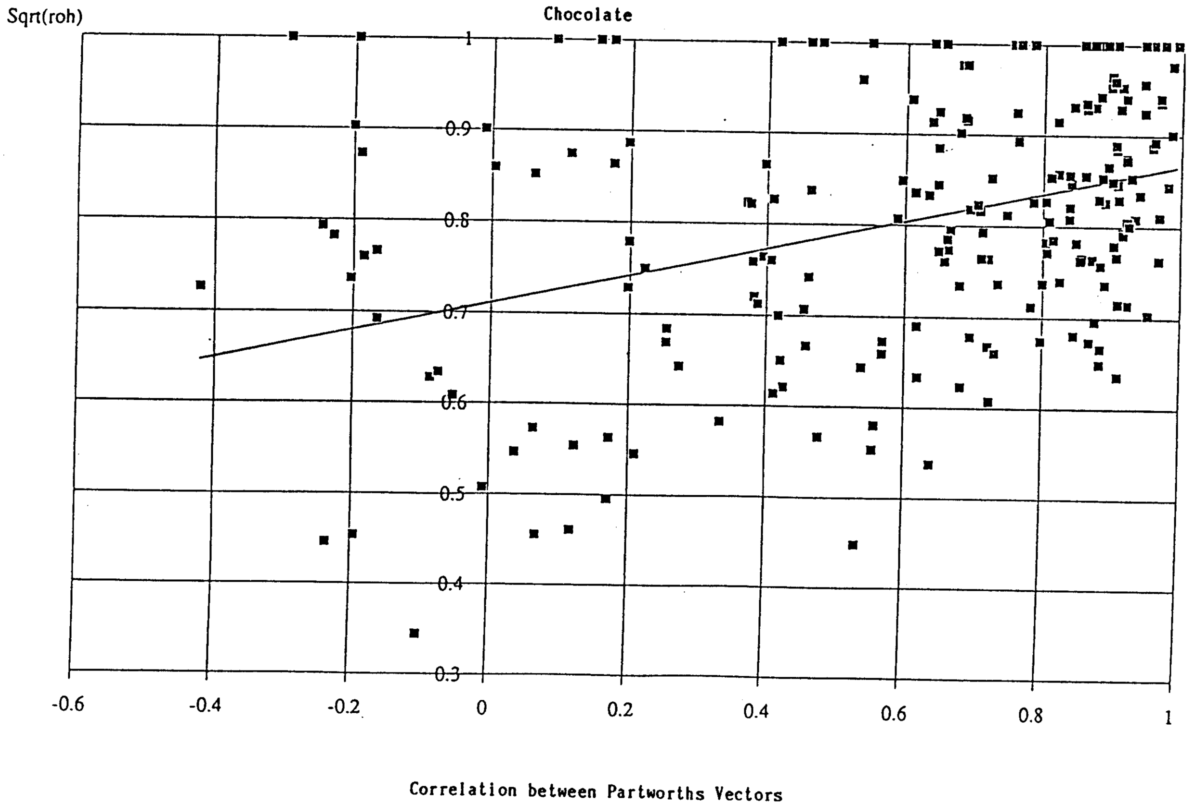


Figure 5

Scatter Plot of Sqrt(roh) v.s. Correlation



more variety-seeked than soft drink.

Table 5
Choice Set Compositions

A. Soft drink

Brand Name	Maker	Category	Size	First Choice
Kirin's Tea	Kirin	Canned Tea	350 ml	79
Coke	Coca Cola	Carbonated	250 ml	64
Five-Mini	Ohtsuka	Isotonic	125 ml	23
Orange-Aide	Suntory	Orange Juice	350 ml	21
UCC Coffee	UCC	Canned Coffee	250 ml	19

B. Chocolate

Brand Name	Maker	Taste	Form	First Choice
m & m	Mar's	Nuts	Flake	62
Hi-Crown	Morinaga	White	Flake	51
Ghana Milk	Lotte	Milk	Plate	47
Black Choco	Meiji	Bitter	Plate	46

Using parameter estimates which they gave in the computer-interactive session just finished, total utility scores could be calculated for all brands in the choice set. Then, the first choice can be predicted and matched with the brand actually picked up by each respondent. "Batting Average" for FP method is 53.9%, which is fairly higher than 44.7% for ACA procedure in case of chocolate. However, ACA with an average of 45.7% could hit the right cans of Cola, Tea, or Orange Juice better than FP only with an average of 40.4%.

The hitting rate of the right brand by full-profile method can be seen in order of a range of likelihood ratio index. Of interest is the fact that the higher

Table 6
Batting Average for Categry by Procedure

		Hit Rate
Chocolate	A C A	44.7%
	C V A	55.3%
	F P	53.9%
Soft drink	A C A	45.7%
	C V A	41.8%
	F P	40.4%

miscalculation of prediction because of absolute lack of degrees of freedom, e.g., fifteen ranked profiles in estimating the nine attribute levels for chocolate.

Figure 6 (B)
Batting Average by roh Coefficient (ACA)

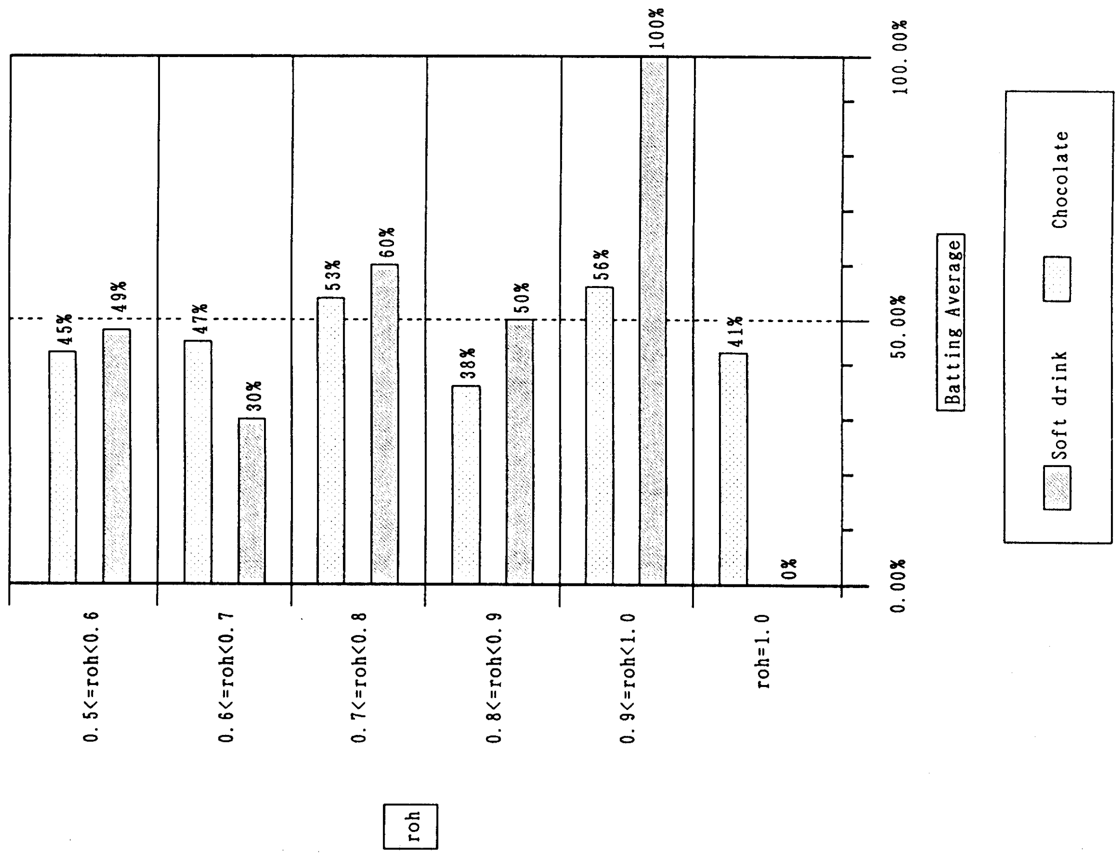
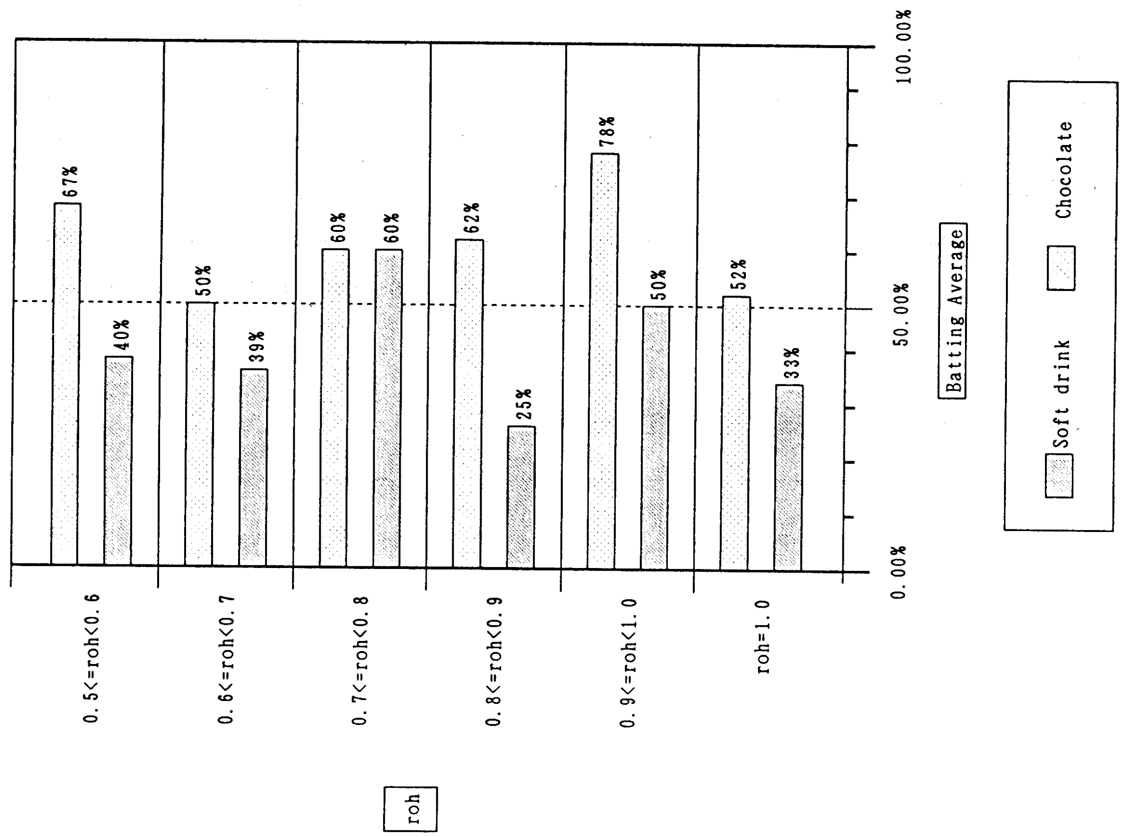


Figure 6 (A)
Batting Average by roh Coefficient (FP)



We recommend that researchers would better make use of ACA against FP, when there are many attributes and/or profiles in a concept evaluation task, since interviewing with ACA is much easier than that with FP. With an increasing number of attributes, accordingly number of product profiles, the time difference in involved between FP and ACA tends to be widen. Care must be taken for prediction when the number of profiles is not so great that an information processing task might not be overloaded.

In summary, it seems to us that the full-profile and the trade-off methods make no significant difference in reliability and predictability.

Other Findings

Finally, we would like to remark two other findings from our conjoint project.

First, relationship of variety-seeking behavior to predictive performance was observed. We asked the respondents for their variety-seeking tendency in their everyday's life as well as in their buying behavior. A variety-seeking measure was scaled as a sum of sensational seeking points, SS Scale named by Kuwabara (1985).[10] The higher SS score a respondent marks, the less predictive respondent's actual choice becomes.

Secondly, as for the usage experience of product category, we could not identify a statistically significant relationship of product experience to external validity. With a larger experience set of brands for soft drink, there is a slightly positive relationship of the consideration set size with "batting average". However, relation is not significant in case of chocolate.

5. Summary and Conclusions

The findings mentioned above are preliminary, since the number of samples in effect was no less than two hundred, but also the respondent's segment was limited to young students and business persons who are accustomed to operating the office automation equipment. Our major findings, however, can be generalized for more applications.

Followings are a summary of findings in this experiment.

First of all, two data-gathering procedures make no difference in estimating the partworths and also in predicting the actual brand choice. In other words, partworths derived by ACA appears very similar with those by Logit or MONANOVA. Secondly, likelihood ratio index can be a good indicator for internal reliability. Thirdly, with an increasing number of attributes and profiles, using PC version of trade-off method is preferable to the conventional ranking method.

Finally, ACA or FP as a market simulator might have an equivalent predictability, although its absolute power must be needed for more extensive research experiments and be applied to a wider variety of product categories.

[**FOOTNOTES**]

(1) This paper is a joint product by Kohsuke Ogawa of Hosei University, Shota Hattori of Kozokeikaku Engineering, Inc., and Tsunehiro Fukushima of Ajinomoto Co., Inc. However, we owe much to our graduate students, Manabu Uchida and Hideyuki Konishi of Hosei University in gathering experimental data from students. One of our staff members from Kozokeikaku, Junichi Kawano, participated in our project to administer the field survey by personal computer. Masahiko Yamanaka of Ajinomoto Co., Inc. should be notified for his contribution. The project was originally proposed by him. He had done a pilot study with first author, and continued to give us valuable comments and suggestions to finish this research project.

(2) The issues regarding data collection methods and the statistical estimation techniques should be referred to Green and Srinivasan (1990).

(3) The 1987 winter issue of <u>Journal of Marketing Research</u> reports in detail two kinds of readily available computer software for conjoint analysis.

(4) Comparison between ACA and FP was discussed in Reibstein et al.(1988) and Green et al.(1988). They could not identify any critical difference in terms of internal validity for both methods.

(5) The profile descriptions here were derived from the tables of 16 trials and 25 trials in Addelman's paper (1962).

(6) The main results in 1990 pilot study was reported in a paper by Uchida(1990).

(7) Partworths here was calculated based on RANK-LOGIT estimation program by Ogawa (1987). To compare with partworths vectors by FP, partworths derived from ACA were rescaled for the norm of a partworths vector to be the number of dimensions. For more detail, see Uchida (1991).

(8) For more detail about Likelihood Ratio Index, see Hauser(1978) or Ogawa(1987).

(9) See Johnson (1987) for reference of Reliability Measure in ACA.

(10) Kuwabara(1985) constructed a Sensational Seeking scale for a proxy variable of variety-seeing tendency. His SS scale was a simple summation of five YES-NO questions.

[**REFERENCES**]

Addelman, S. (1962), "Orthogonal Main Effect Plans for Asymmetrical Factorial Experiments," <u>Technometrics</u>, <u>4</u> (February), 21-46.

Green, P.E., K. Helsen, and B. Shandler (1988), "Conjoint Internal Validity under Alternative Profile Presentation," <u>Journal of Consumer Research</u>, <u>15</u> (December), 392-397.

Green, P.E. and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," <u>Journal of Consumer Research</u>, <u>5</u> (September), 103-123.

Green, P.E. and V. Srinivasan (1990), "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," Journal of Marketing, 64 (October), 3-19.

Johnson, R.M. (1974), "Trade-off Analysis of Conjoint Values," Journal of Marketing Research, 11 (May), 121-127.

Johnson, R.M. (1987), "Adaptive Conjoint Analysis," in Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing, Ketchum, ID: Sawtooth Software, 253-256.

Kruscal, J.B. (1965), "Analysis of Factorial Experiments by Estimating Monotone Transformations of Data," Journal of the Royal Statistical Society, Series B, 27, 251-263.

Kuwabara, T. (1985) "Life Style Analysis in view of Consumer Sensation Seeking," Shohi-to-Ryutu, 9, 156-165.

Ogawa, K. (1987) ," An Approach to Simultaneous Estimation and Segmentation in Conjoint Analysis," Marketing Science, 6 (Winter), 66-81.

Reibstein, D., J.E. John, and W. Boulding, (1988) ,"Conjoint Analysis Reliability: Empirical Findings," Marketing Science, 7 (Summer), 271-285.

Uchida, M. (1991), "An Empirical Research in Testing Reliability and Comparing Predictive Validity of Conjoint Measurements," Master Theses, Graduate School of Economics, Hosei University.

Wittink, D.R. and P. Cattin (1989) ,"Commercial Use of Conjoint Analysis: An Update," Journal of Marketing, 53 (July), 91-96.