AN EMPIRICAL EXAMINATION OF PREDICTIVE VALIDITY
OF ALTERNATIVE CONJOINT DATA COLLECTION PROCEDURES

Madhav N. Segal, Southern Illinois University - Edwardsville
Roger H. Gates, University of Texas at Arlington

Abstract

This empirical investigation examines variations in measures of predictive validity of conjoint measurement solutions as a result of using different data collection procedures. The two most commonly used conjoint methods are comparatively evaluated: Multiple-Factor-Evaluation (MFE) and Two-Factor-Evaluation (TFE). The results show on an average basis the validity measures reported for both procedures are extremely good.

Introduction

The model of conjoint measurement (CM) which has been used to estimate trade-off utilities and quantify judgemental type data, has generated much interest and concern in the field of marketing research in the past few years. The applications of CM to problems and situations in marketing have been illustrated by Green and Rao (1971) amongst others.

Conjoint measurement has been validated as a useful technique for modelling consumer preferences and in predicting their behavior toward new stimuli. However, much empirical work remains to be done to settle several methodological issues and provide concrete guidelines for applying CM effectively to different marketing problems. As of today, conjoint analysis is far from being a settled, cut-and-dried methodology (Green and Srinivasan 1978).

Conjoint measurement has been applied to marketing research problems in several different ways. However, the major differences arise principally from the two methods that are utilized for collecting conjoint analysis data from consumers (Green and Srinivasan 1978). The two data collection methods are known as two-factor evaluation (TFE) and multiple-factor-evaluation (MFE) techniques. Thus TFE and MFE data collection methods can be looked at as two different measurement instruments, each trying to accomplish the same purpose. To this date there is a little research documentation in existence to show whether or not these two alternative data collection methods lead to different conjoint solutions. Moreover, issues of validity and reliability have not been treated comprehensively in any of the research papers dealing with this subject matter. The main purpose of this study is to document the existence, magnitude, and distribution of inter-method differences/similarities in predictive efficiency relating to the MFE and TFE conjoint data collection procedure.

Methodology: Research Question and Study Design

Research Question

Conjoint measurement studies are generally carried out using hypothetical product profiles. The external validity can be tested by comparing predictions against a respondent's actual behavior with respect to real stimuli. Thus external validity deals with issues of reproducibility and predictability of external criteria. This method is also referred to as predictive validity by Parker and Srinivasan (1976, p. 101). The research issue addressed in this empirical investigation deals with the determination of differences in measures of predictive validity between the MFE and TFE data collection procedures. More specifically, the research question can be stated as follows:

RQ: Does either data collection procedure produce part-worth scales which provide superior predictions of appropriate external stimuli?

Experimental Design

A convenience sample of approximately 50 undergraduate students was selected for empirical study. None of these respondents had any prior experience with nonmetric multivariate techniques in marketing research. The respondents were told that they were participating in a marketing experiment and that they should use their own preferences to evaluate various product concepts. To maintain a reasonable level of interest for most respondents and to enhance the validity of the study, the choice situation selected was consumer preferences for apartments to rent. Student involvement with such a problem situation was believed to have provided the respondents with a relevant framework for identifying preferences.

Three attributes were chosen to constitute apartment alternatives. These were number of bedrooms, rent per month, and distance to campus. Table 1 displays the attributes and levels used. These attributes and levels were chosen to represent salience and range of apartment attributes available in the study area at the time the research was conducted (Darrell 1979).

TABLE 1

APARTMENT ATTRIBUTES AND LEVELS
USED IN THE RESEARCH DESIGN

Attribute		Level
A - Number of Bedrooms	A_1:	Efficiency
	A_2:	One Bedroom
	A_3:	Two Bedroom
B - Rent Per Month	B_1:	\$150
	B_2:	\$200
	B_3:	\$250
C - Distance to Campus	C_1:	Walking
	C_2:	Bicycle
	C_3:	Driving

The three factorial experimental design was used to maintain simplicity in administration and data analysis. More complex apartment profiles with additional attributes could be used. However, since the main focus of the research study is to make comparative analysis between the two data collection procedures, it was deemed unnecessary to introduce confounding effects in the experimental design (such as using fractional factorial experimental designs because of using several attributes). Additionally, three levels for each attribute were selected, thus, giving a symmetric and comparable design from the standpoint of data-collection instrument development.

The TFE data collection instruments were developed using a full design where all attributes were considered in pairs. Therefore, the total number of evaluations for the three trade-off grids was 27. For the MFE procedure, a full factorial design using all three attributes at three levels each was used to generate 27 (3x3x3) combinations of hypothetical apartment alternatives. Nine validation profiles were constructed by randomly selecting them from the original set of 27 apartment concept profiles.

Experimental Tasks: Data Collection Procedure

In the first phase of the experiment, each respondent supplied rank ordered preference for all pairs of the three apartment attributes. This resulted in 27 evaluative judgements for the three trade-off grids. All these respondents also provided data for the MFE procedure. The respondents evaluating the MFE concepts were instructed to sort the profiles (given on a 3x5 card) into four categories: excellent, good, fair, and poor housing deal for the money. After performing this part of the task, each subject ordered the cards within each category from best to worst. This sequential procedure resulted in a strict rank order of the 27 cards from best to worst housing deal for the money for each subject. Both the MFE and the TFE tasks were randomized prior to presentation so as to avoid any order bias. All respondents performed an intervening task (filling out a short questionnaire on background information) prior to evaluating the validation profile-set.

Data Analysis

The theory of conjoint measurement applies to ordinal data structures and in a typical application a subject is asked to rank order a set of multiattribute profiles by a specified criterion, such as perceived utility, thus quantifying his subjective judgements of trade-offs between attribute levels. The general main-effects additive multiattribute model is hypothesised to apply to the problem situation and can be formulated as (Barron 1977):

$$\phi(Y) = \alpha(A) + \beta(B) + \gamma(C)$$

where,

$\phi(Y)$ = an order preserving (monotone) function of the observed response (rank order Y)

$\alpha(A)$ = a part-worth function defined over attribute A (bedrooms)

$\beta(B)$ = a part-worth function defined over attribute B (rent)

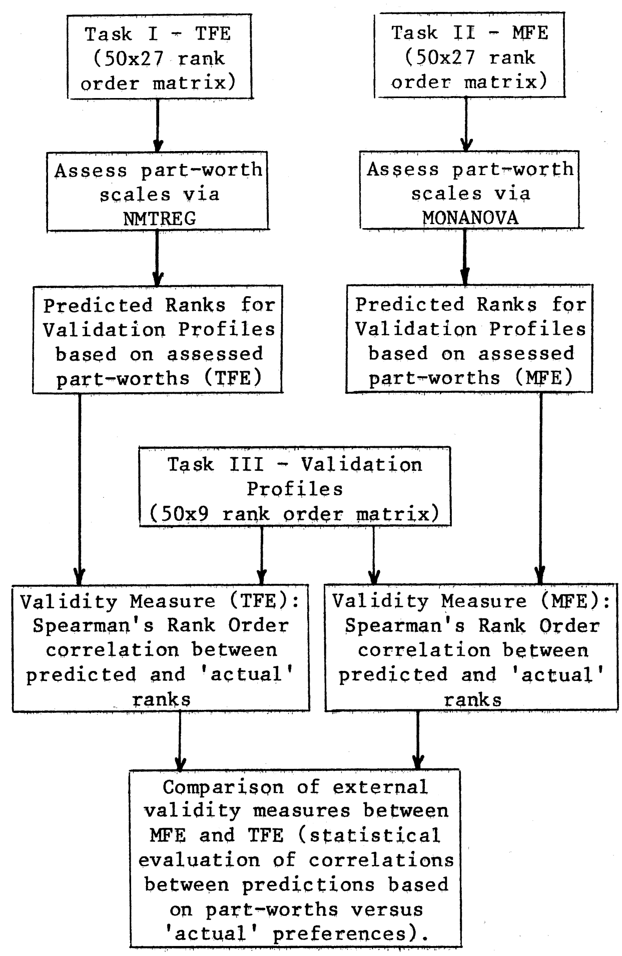
$\gamma(C)$ = a part-worth function defined over attribute C (distance)

The estimation problem with the TFE procedure ordinarily consists of finding a set of individual part-worths so that their pairwise sums have desired rank orders. In comparison, the MFE procedure involves determination of part-worths for each attribute which best accounts for the rank orders assigned to the overall concepts evaluated by the respondents. In general, the research interest is in computing numerical scale values for the dependent and the independent variables regardless of the approach that is used for collecting the ordinal data.

An overview of the basic steps involved in data analysis is given in Figure 1. Because of incomplete

Figure 1

FLOW CHART PLAN OF DATA ANALYSIS*



*(All analysis done at the individual subject level)

response to either the MFE or TFE tasks, only 44 responses were usable for the complete analysis. This constituted an 88% response rate which was about what was expected. A brief summary of each step for data analysis follows:

Data set I included respondent evaluations of three trade-off matrices (a 50x27 rank order matrix). Johnson's Nonmetric Regression (Johnson 1973) program was used to assess part-worth scales for each subject. Kruskal's Monotone Analysis of Variance (1965) program was applied to each subject's rankings of the 27 apartment profiles (Data set II - a 50x27 rank order matrix) and part-worth scales were assessed for each component. Data set III included respondent supplied nine preference rankings for the validation profiles (a 50x9 rank order matrix).

Results

This section presents and examines the results on the ability of alternative data collection methods to provide predictions of appropriate external criteria. The results on the predictive efficiency of the MFE and TFE procedures are presented first and then a comparative analysis is made on the obtained results.

Predictive Validity of MFE Procedure

Using the part-worth preference functions from MONANOVA, the overall utility for the nine validation profiles according to the additive model were estimated and then preference rankings were assigned. This resulted in nine predicted rankings of most preferred to least preferred apartments. These predicted rankings were then correlated with the actual rankings for nine validation profiles obtained from respondents. Spearman's rank order correlation was used as a measure of association between the predicted and actual rankings.

Table 2

FREQUENCY TABULATION OF PREDICTIVE VALIDITY MEASURES (SPEARMAN'S CORRELATION COEFFICIENT BETWEEN ACTUAL AND PREDICTED RANKS) FOR THE MFE PROCEDURE (n=43)

Frequency Tabulation:

Correlation Interval	Frequency	Cum Frequency	Percent	Cum Percent
$0.95 < r \leq 1.00$	25	25	58.14	58.14
$0.90 < r \leq 0.95$	6	31	13.95	72.09
$0.85 < r \leq 0.90$	7	38	16.27	88.36
$0.80 < r \leq 0.85$	3	41	6.97	95.33
$0.75 < r \leq 0.80$	0	41	0.00	95.33
$0.70 < r \leq 0.75$	2	43	4.65	100.00

Table 3

FREQUENCY TABULATION OF PREDICTIVE VALIDITY MEASURES (SPEARMAN'S CORRELATION COEFFICIENT BETWEEN ACTUAL AND PREDICTED RANKS) FOR THE TFE PROCEDURE (n=43)

Frequency Tabulation:

Correlation Interval	Frequency	Cum Frequency	Percent	Cum Percent
$0.95 < r \leq 1.00$	16	16	37.20	37.20
$0.90 < r \leq 0.95$	8	24	18.60	55.80
$0.85 < r \leq 0.90$	6	30	13.95	69.75
$0.80 < r \leq 0.85$	4	34	9.30	79.05
$0.75 < r \leq 0.80$	2	36	4.65	83.70
$0.70 < r \leq 0.75$	2	38	4.65	88.35
$r \leq 0.70$	5	43	11.62	100.00

Table 2 illustrates the frequency tabulation and summary measures of predictive validity (correlations between the predicted and actual ranks) results for the MFE data collection procedure. These results are shown for 43 respondents after deleting one outlier whose coefficient value was far out of line with the remaining group. Data reported in the table are indicative of a very high level of validity for the MFE data collection procedure. Without any exception, all respondents have a correlation coefficient of 0.7 and higher. Also, surprisingly for nearly 60% of all respondents correlations of greater than 0.95 are reported.

There were 12 respondents whose predicted ranks matched perfectly with actual ranks and therefore reported a coefficient of exactly equal to 1.00. All of these coefficients were highly significant at 0.05 level and beyond. The mean correlation coefficient ($\bar{r}$) has a value of 0.945 which indicates that on an average basis, the MFE procedure's ability to predict actual behavior is extremely good.

Predictive Validity of TFE Procedure

Utilizing the part-worth functions (from NMTREG), the utility for the nine validation profiles according to the additive model were estimated and then preference rankings were assigned. These predicted ranks from the TFE procedure were then correlated with the actual rankings for nine validation profiles. This resulted in nine predicted rankings of the most preferred to least preferred apartments for each respondent. Similar to the MFE procedure, Spearman's correlations were computed between the TFE predicted ranks and the actual ranks. These coefficients provided measures of predictive validity for the TFE data collection procedure.

Table 3 illustrates the frequency tabulation and summary measures of predictive validity (correlation between the predicted and actual ranks) results for the TFE data collection procedure. Data reported in the table are indicative of relatively good validity results for the TFE procedure. Approximately 88% of all respondents have a correlation value of 0.7 and higher. Also, a correlation of 0.9 and higher is reported for 24 respondents (nearly 56%). There were 8 respondents whose predicted ranks matched perfectly with actual ranks and therefore, reported a coefficient of exactly equal to 1.00. There were five respondents which had a coefficient value of 0.7 and less. The lowest value of the coefficient was equal to 0.23. With an exception of three, all other coefficients were significant at 0.05 level and beyond. The mean correlation coefficient ($\bar{r}$) has a value of 0.874 which indicates that on an average basis the TFE procedure's ability to predict actual behavior is rather good.

Comparative Analysis of Inter-Method Differences in Validity

Even though no specific null and alternative hypotheses were formulated in this area, one can statistically evaluate the differences in validity measures of MFE and TFE data collection procedures. To test if the differences in validity measures on an average basis were significant, a matched-pair t-test was utilized.[1] The difference in correlation coefficients (d_r) was calculated for each respondent by using the equation: d_r = (correlation between MFE predicted ranks and actual ranks) - (Correlation between TFE predicted ranks and actual ranks). The mean difference ($\bar{d}_r$) is considered to be the basis of this test. Thus, it is hypothesized that the mean difference $\bar{d}_r = 0$, meaning there are no differences in average validity measures reported for the MFE and TFE data collection methods.

[1] The sample mean $\bar{d}_r$ is assumed to be normally distributed (invoking the central limit theorem) with a mean of 0.0 and a standard deviation of $\sigma d/\sqrt{n}$. The test statistic is t* where:

$$t^* = \frac{\bar{d} - 0}{\sigma d / \sqrt{n}}$$

is distributed as t with (n-1) degrees of freedom. The decision rule for assuming an alpha risk of 0.05 will be: if $t^* > t_{n-1, 0.05}$, then reject the null hypothesis.

All relevant statistics to carry out a paired t-test and the results of the test are displayed in table 4. The mean difference ($\bar{d}_r$) between MFE-correlation and TFE-correlation coefficient is 0.071 with a standard deviation of 0.145. The calculated t value of 3.21 is greater than the critical t values for both assumed levels of alpha risk (0.01 and 0.05). The null hypothesis is therefore, rejected at both 0.01 and 0.05 alpha risk levels.

Table 4

SUMMARY COMPARISON OF DIFFERENCES IN MEASURES OF PREDICTIVE VALIDITY (CORRELATION BETWEEN PREDICTED AND ACTUAL RANKS) AND PAIRED t-TEST FOR DIFFERENCES BETWEEN MFE-CORRELATION AND TFE-CORRELATION (n=43)

Variable*	Mean	Standard Error of Mean	t-Value	Significance** α=0.01	Significance** α=0.05
r(MFE)	0.94	0.01	NA	NA	NA
r(TFE)	0.87	0.02	NA	NA	NA
d_r	0.07	0.02	3.21	sig.	sig.

*r(MFE) = rank correlation between predicted and actual ranks (MFE)
*r(TFE) = rank correlation between predicted and actual ranks (TFE)
dr = difference between r (MFE) and r (TFE)

**Significance is tested with the following t critical values:

t_c (α = 0.01 and 42 d.f.) = 2.704

t_c (α = 0.05 and 42 d.f.) = 2.021

Thus, test indicates that there is a significant difference between the validity measures for the MFE and TFE data collection procedures with a favor for the MFE procedure. However, it is worth noting that even though differences in predictive (external) validity were found to be significant, the average validity measures were quite good for both data collection procedures.

The major findings of the study can be reported as: (a) The average correlation between the predicted and actual ranks for the MFE procedure was equal to 0.945. Therefore, on an average basis the validity measures reported for the MFE procedure were extremely good. (b) The average correlation between the predicted and the actual ranks for the TFE procedure was equal to 0.874. Therefore, on an average basis the validity measures reported for the TFE procedure were very good. (c) There is a significant difference between the validity measures for the MFE and TFE data collection procedures with a favor for the MFE procedure.

Conclusion

The validity results for the two data collection procedures indicate that each procedure yields predictions which compare extremely well with the actual judgments. This is noteworthy in light of the fact that only ranks were employed as input to the models. The results obtained are in conflict with the findings reported by Montgomery, Wittink, and Glaze (1977) who, in a study of job choice by MBAs, found that the TFE procedure yielded higher predictive validity than the MFE procedure. However, these findings support the results on the validity issue reported by Rao and Solgaard (1978) in the context of tenure decision recommendations for educators. Also, research findings of this study are in congruence with the ones recently reported by Jain et al. (1979) who found that the data type does not affect the predictive efficiency in the context of a bank selection.

Therefore, it is concluded that on an average basis the validity measures reported for both procedures were extremely good. However, on evaluating the differences statistically, the MFE data collection procedure was found to produce results which provided superior predictions of appropriate apartment profiles. For most applied market research studies this difference in predictive power should not be a cause for alarm because even the TFE procedure produces results which provide very good predictions.

Even though this study suggests comparable predictive powers for the two most commonly used conjoint data collection procedures, such a conclusion must be tempered with caution. There are two specific reasons for this. First, no actual selections of apartments were available and validation profiles therefore, were hypothetical product profiles. Second, the validation profile set presented to the respondents resembled a typical MFE profile set and this possibly may have biased the results for the MFE procedure.

Therefore, it is suggested that both the TFE and MFE data collection procedures are equivalent in terms of producing results which provide very good predictions. However, the issue of which data collection procedure is superior cannot be answered till other bases of comparisons, such as stability of part-worths, respondent time and associated interviewing costs, etc. are examined and evaluated empirically (Segal 1979).

References

Hutton F. Barron (1977), "Axiomatic Conjoint Measurement", Decision Sciences, 8, 48-59.

Dunn Darrell (1979), "Student Housing", Shorthorn, University of Texas--Arlington Student Newspaper, 1-2.

Paul E. Green and V.R. Rao (1971), "Conjoint Measurement for Quantifying Judgemental Data", Journal of Marketing Research, 8, 355-63.

Paul E. Green and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook", The Journal of Consumer Research, 5 (September), 103-23.

Arun K. Jain, Franklin Acito, Naresh K. Malhotra, and Vijay Mahajan (1979), "A Comparison of Internal Validity of Alternative Parameter Estimation Methods in decompositional Multiattribute Preference Models", Journal of Marketing Research, Vol XVI, (August), 313-22.

Richard M. Johnson (1975), "A Simple Method for Pairwise Monotone Regression", Psychometrika, 40 (June), 163-8.

Richard M. Johnson (1973), "Pairwise Nonmetric Multidimensional Scaling", Psychometrika, Vol. 38, No. 1, pp. 11-8.

Joseph B. Kruskal (1965), "Analysis of Factorial Experiments by Estimating Monotone Transformations of the

Data", Journal of the Royal Statistical Society, Series B (March), 251-63.

David B. Montgomery, Dick R. Wittink, and Thomas Glaze (1977), "A Predictive Test of Individual Level Concept Evaluation and Trade-off Analysis", Research Paper #415, Graduate School of Business, Stanford University.

Barnett R. Parker and V. Srinivasan (1976), "A Consumer Preference Approach to the Planning of Rural Primary Health Care Facilities", Operations Research, 24, 995-1,025.

Vithala R. Rao and Hans S. Solgaard (1978), "An Empirical Evaluation of Alternative Multiattribute Utility Models", Research Frontiers in Marketing: Dialogues and Directions, (1978 Educators' Proceedings), Series #43, Chicago: American Marketing Association.

Madhav N. Segal (1979), "Variation Among Conjoint Measurement Solutions: An Empirical Examination of the Effect of Data Collection Differences", Unpublished Doctoral Dissertation, University of Texas at Arlington.