INCONSISTENCY RATIO AND PREDICTION: SOME EMPIRICAL FINDINGS.

Rajendra Mulye
Faculty of Business, Marketing Group
Royal Melbourne Institute of Technology, GPO Box 2476 V, Melbourne, Vic 3001 Australia
rajum@rmit.edu.au

Abstract: The role of inconsistency ratio in improving predictive validity of the Analytic Hierarchy Process is examined in an empirical setting. Evaluation of running shoes is used as the decision task. Inconsistency ratios generated at two levels of the hierarchy are compared individually and simultaneously in terms of their ability to predict ratings of an external holdout sample. The results indicate a weak but significant relationship between the inconsistency ratio and the model's prediction.

Introduction

The data for this study comes from a larger study that compared AHP with conjoint analysis, details of which are published in Mulye (1995). Since the intent of the original study was to compare two widely different methods of attribute valuation, the effect of inconsistency ratio on judgement accuracy was not addressed. The purpose of this paper is to address this issue. Specifically, the relationship between the inconsistency ratio and performance of the AHP on the criterion of predictive validity is examined. This is an important issue because respondents are required to carry out a complete set of n(n-1)/2 comparisons instead of (n-1) comparisons to obtain the inconsistency ratio and the potential benefit of these additional comparisons in improving prediction needs to be verified. This is particularly critical in marketing applications of the type described in this paper where respondent participation, interview time, and reliability issues relating to respondent fatigue are a major consideration. If the inconsistency ratio proves to have marginal influence on the predictive performance of the AHP then its inclusion in applications of attribute valuation should be weighted against the potential benefits of incorporating a large number of attributes and reducing interview time. The original study is described first followed by analysis of data pertinent to the inconsistency ratio and discussion of results.

Original Study

The objective of this study (Mulye, 1995) was to compare two maximally dissimilar methods of attribute valuation in an empirical setting. Such studies help to understand whether any given approach is better in a problem situation and whether one evaluation method will lead to the same conclusion as the other. The AHP and Conjoint analysis method were used in this study since they are least alike on several characteristics such as scaling, preference elicitation, synthesis, etc which are summarised in Mulye (1995). Both techniques are based on the concept of trade-off to derive weights for a multiattribute product and its attributes under a common model and are therefore directly comparable.

The Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) proposed by Saaty (1980) can be classified under the category of compositional approach where a respondents' preference for a multiattribute product is derived as a weighted sum of its perceived attribute levels, with the weights represented as importance ratings of the products corresponding attributes. In the AHP, a multiattribute problem is first structured in to a hierarchy of interrelated elements and then pairwise comparison of elements in terms of their dominance is elicited. The weights are given by the eigen vector associated with the highest eigen value of the reciprocal ratio matrix of pairwise comparisons. The process also generates an inconsistency ratio which is claimed to guide respondent in their evaluation task. The AHP can be represented by the following equation:

$$V_{ik} = \Sigma\, w_j \; x_{jk} \qquad (1)$$

where
V_{ik} = overall evaluation of alternative k by individual i.
w_j = importance weights given to attribute j.
x_{jk} = extent to which attribute j is offered by alternative k.

Conjoint Analysis

In contrast to the compositional approach of AHP, conjoint analysis falls under the category of decompositional approach where a consumer's total preference for a multiattribute product is used to infer a consumers' underlying cognitive structure. In conjoint analysis a consumers' response to a set of product profiles is used to derive partworths for its attributes assuming some form of preference model. When the preference model is assumed to be linear additive, the conjoint analysis procedure is directly comparable to the AHP approach. Conjoint analysis can also be represented by the above equation when the x_{jk} represents the presence or absence of an attribute level.

Methodology

A convenience sample of 120 undergraduate business students at Simon Fraser University was used in this study. The subjects were paid five dollars as an incentive to participate and provide careful estimates. The task was evaluation of running shoes used for every day wear. Running shoes was considered a suitable stimulus because of its familiarity and relevance to the fashion and fitness conscious student population. Four critical attributes given in table 1 below were selected on the basis of focus group and indepth interview with students and shoe retailers respectively.

Table 1 Critical attributes of a running shoe

ATTRIBUTE	*ATTRIBUTE LEVELS*			
Color	White	Black	Grey	
Brand	Nike	Reebok	Converse	Brooks
Style	High tops	Mid cut	Low cut	
Price	$ 40	$ 60	$ 80	$ 100

In the conjoint task, sixteen profiles were generated by an asymmetric orthogonal array plan. In addition, five hold out cards were generated to test for predictive validity. Two measurement scales were used to evaluate the profiles - a 10 point rating scale and an ordinal scale. Standard instruction of ordering and rating the profiles was used.

The AHP task was completely computerised through the use of a modified version of Fuzzy Choice (Fuzzy Choice, 1987) software program. The subjects were familiarised with the program through a trial run of evaluating and selecting the best car from amongst four cars described on two attributes. The subjects were randomly assigned to one of the following six cells (table 2) of an order controlled within-subject design. Two other versions of the AHP suggested by Schoner and Wedley (1989) - referenced AHP and B-G modified AHP were also used. The methods differed only in their questioning of attribute importance. In the referenced AHP approach the question of attribute importance was made explicit to reflect the normalisation process and respondents were asked to compare the average of the alternatives under each criterion. In the B-G modified approach respondents compared the attributes with respect to the highest weighted alternative, however the normalisation process was identical to that of conventional AHP. The absolute measurement approach suggested by Saaty (1987) was used for all the three AHP methods.

Table 2 Experimental design for study 1.

METHODS	*Conjoint Rating*	*Conjoint Ranking*
Conventional AHP	Cell 1 (n=20)	Cell 4 (n=20)
referenced AHP	Cell 2 (n=20)	Cell 5 (n=20)
B-G AHP	Cell 3 (n=20)	Cell 6 (n=20)

Each subject performed two evaluations using conjoint and the AHP approach. To account for order effects, half of the group in each cell performed the evaluations in reverse order. The holdout cards were always administered at the end of the interview. The conjoint partworths and the AHP derived weights of the attribute levels were used to score the five holdout profiles for each method. These scores were then correlated with the stated ranking or rating of the holdout profiles to access predictive validity of the methods.

The main finding of this study is that there is no significant difference in prediction among the three AHP methods. The questioning methods had little bearing on how the respondents interpreted the question of attribute importance, although respondents seem to have responded in a consistent manner to this question. It is therefore appropriate to pool the data for further analysis to verify the hypothesis of the current study as to whether the inconsistency ratio has any influence on the prediction correlation.

Inconsistency ratio

One of the most important aspects of the AHP is considered to be its ability to provide a measure of overall consistency of judgement throughout the evaluation process. The measure is calculated from the principle eigen value λ of the matrix of pairwise comparison. Saaty (1980) has shown that when a matrix is perfectly consistent the principle eigen value is equal to the order of that matrix. When it is inconsistent the principle eigen value exceeds n. Thus, the departure of λ from n is used as a measure of inconsistency by taking a ratio of this difference to the average of the corresponding difference of eigen values of a large number of matrices of randomly generated comparisons. The consistency index is given by λ-n/n-1 and the inconsistency ratio is given by (CI * 100)/ACI where ACI is the average index of randomly generated weights. Higher values of this ratio indicate increased inconsistency. As a rule of thumb a ratio of 10% or less is considered good and less than 20% considered acceptable otherwise it is recommended to change some of the comparisons to achieve a better consistency. However, consistency is not synonymous with accuracy of judgement. It is possible for a person to be perfectly consistent and inaccurate at the same time. For example, a person evaluating heights of three buildings which are 100m, 200m and 400m high (in the ratio 1:2:4) can estimate them incorrectly but consistently as follows (in the ratio 1:2:6):

1	1/2	1/6
2	1	1/3
6	3	1

Also note that technically one can derive weights for n items with only (n-1) comparison; if any one weight is assigned a value of 1, then ratios with the other attributes give their relative importance. In contrast, the AHP requires n(n-1)/2 comparisons. The redundant comparisons are used to calculate the inconsistency ratio and to average out judgement errors in the final weights which is considered to be an important aspect of the AHP. This however poses practical problems for large size problems where the evaluation task can get unwieldy with increase in the number of elements in a level or the number of levels in a hierarchy. Even for a small size problem such as the one used in this study, with only four attributes a total of 24 comparisons and an average of 30 minutes was required to complete the task including the time required to familiarise the respondents with the AHP task.

Inconsistency data

The data were analysed to verify the hypothesis that: the inconsistency ratio is inversely proportional to the holdout prediction correlation ie. whether or not subjects with lower inconsistencies in their paired judgement tended to have better predictive power and vice versa. Since a total of five inconsistency ratios was generated, four at the bottom level of the hierarchy for attribute level comparison and one at the next higher level for attribute comparison, the following four measures were used to incorporate them in the analysis.

1) The correlation between the holdout prediction correlation and the average of the four inconsistency ratios generated in comparison of the attribute levels at the bottom level of the hierarchy.

2) The correlation between the holdout prediction correlation and the inconsistency ratio generated in comparison of the attributes at the second level of the hierarchy,

3) The correlation between the holdout prediction correlation and a weighted average of the inconsistency ratio. A weight of four is assigned to the inconsistency ratio at the attribute level and one each for the four attribute level comparison ratios. A larger weight is assigned to the inconsistency ratio generated at the second level of the hierarchy because the methods differed in their questioning procedure at this level and it was believed that such differences would be manifested in the composite weight and consequently in calculating the overall weight of the holdout profiles. The questioning procedure at the bottom level of the hierarchy was identical for all the AHP methods and therefore a weight of one was assigned to the inconsistency ratios generated at this level.

4) The simultaneous and individual effect of all the five inconsistency ratio on the holdout prediction correlation. This measure was based on regression analysis. The holdout prediction correlation was regressed on the five inconsistency ratios to test the individual and joint effect of the inconsistency ratios.

Results

The Pearson's correlation for the first three measures is -0.266, -0.208, and -0.263 respectively. All the correlations are significant at the .05 level and in the expected direction. However, the magnitude of the correlations is lower than expected and consistent across the three measures.

The regression results of the fourth measure are given in table 3. The results for this measure are not different from the first three measures. The R square of .1189 suggests a very weak relationship between the inconsistency ratios and the predictive validity of the models. The regression coefficients are in the expected direction and are all insignificant except for the inconsistency ratio for the attribute colour. While there is no clear explanation for this finding, one possible explanation could be from the observation that the attribute colour was also given the highest overall weight by the respondents and thus the respondents may have been more confident (and therefore more consistent) in their evaluation of this attribute.

Table 3 Regression report of inconsistency ratios on holdout prediction correlation.

		Inconsistency Ratios				
	Constant	Colour	Brand	Height	Price	Attribute
reg. coeff.	0.8056	-0.7964	0.1939	-0.00391	-0.2833	-0.27575
t value		-2.79	0.29	-0.03	-0.87	-1.12
p value		0.006	0.774	0.978	0.385	0.264

Conclusion

The results do not indicate a strong support for the hypothesised relationship between the inconsistency ratio and the holdout prediction correlation. Although, the correlations are in the expected direction, their magnitude is too low to make any definite conclusion on the influence of the inconsistency ratio in improving AHP predictions. It seems momentous to consider approaches suggested for reducing the number of comparisons in applications where respondent fatigue is a major consideration. Weiss and Rao (1987) propose using different decision makers to complete subsets of attributes at each level in a balanced incomplete block design. This is similar to the approach suggested in the conjoint literature of allocating fraction of a set of profiles from a factorial design amongst different groups of respondents. A study that compares the reliability of these two approaches could be envisaged. Another approach to reducing the number of pairwise comparison is the method of incomplete pairwise comparison developed by Harker (1987) in which questions at a node are ordered in decreasing information value and the comparison process is stopped when added value of questions decreases beyond a prespecified level. The process starts with only one redundant comparison at each node to enable the calculation of the inconsistency ratio. If the ratio is too high the comparison process is continued. Two stopping rules are proposed. The first stopping rule states that comparisons at a node should stop if the marginal difference in the weights of the local priorities from asking an additional question is less than some prespecified value. The second stopping rule states that an additional question will be required only if it contributes to rank reversals in the local priority weights.

The usefulness of the inconsistency ratio, in an absolute sense, for other decision problems cannot be made from these findings. This is because respondents with high inconsistency ratio had the opportunity to go back and alter some of the comparisons to obtain a lower inconsistency ratio. Also no record was kept of respondents who attempted to use the inconsistency ratio as a guide in their comparison, the level of inconsistency ratio they were willing to accept, the time taken to complete the evaluation process and the general satisfaction with the AHP output. A controlled experiment that overcomes the above limitation is required to further validate the link between the inconsistency ratio and predictive power of AHP.

References

1. Fuzzy Choice (1987) *General Decision Support Systems Inc.*

2. Harker, P.T (1987) "Incomplete Pairwise Comparisons in the Analytic Hierarchy Process"", *Mathematical Modelling*, 9, 353-360.

3. Mulye, R. (1996) "An Empirical Comparison of the Analytic Hierarchy Process and Conjoint Analysis", *Proceedings of the 1996 Australian Marketing Educators' Conference*, Adelaide, 399-419.

4. Saaty, T.L (1980) *The Analytic Hierarchy Process*, New York: Mc-Graw Hill.

5. Saaty, T.L. (1987) "Rank Generation, Preservation and Reversal in the Analytic Hierarchy Process", *Decision Sciences* 18 157-177.

6. Schoner B. and Wedley W. (1989) "Ambiguous Criteria Weights in AHP: Consequences and Solutions", *Decision Sciences*, 20, 462-475.

7. Weiss, E.N. and Rao, V.R. (1987) "AHP Design Issues for Large Scale Systems", *Decision Sciences*, 18, 43-61.