PAUL E. GREEN and KRISTIAAN HELSEN*

Recently, both Hagerty and Kamakura have proposed insightful suggestions for improving the predictive accuracy of conjoint analysis via various types of averaging of individual responses. Hagerty uses Q-type factor analysis (i.e., optimal weighting) and Kamakura a hierarchical cluster analysis that optimizes predictive validity. Both approaches are compared with conventional conjoint and self-explicated utility models using real datasets. Neither the Hagerty nor the Kamakura suggestions lead to higher predictive validities than are obtained by conventional conjoint analysis applied to individual response data.

Cross-Validation Assessment of Alternatives to Individual-Level Conjoint Analysis: A Case Study

Improving the predictive accuracy of conjoint analysis has been an important research objective (Green and Srinivasan 1978) since the technique's introduction and application to marketing problems. Early attempts to improve predictive validity by full or partial response aggregation across respondents include componential segmentation (Green 1977; Green and DeSarbo 1979), quantal choice models, such as the logit and probit (Currim 1981; Madansky 1980; McFadden 1980), hybrid conjoint models (Green, Goldberg, and Montemayor 1981), Bayesian models (Cattin, Gelfand, and Danes 1983), and empirical Bayes' updating (Green and Krieger 1989).

However, as Moore (1980) and Srinivasan (1980) have pointed out, data aggregation does not necessarily lead to higher cross validation. In fact, the traditional full profile conjoint analysis has held up well in comparison with estimation procedures that make use of various ways of aggregating response data (see Akaah and Korgaonkar 1983 and Green 1984 for comparisons).

*Paul E. Green is the S. S. Kresge Professor of Marketing and Kristiaan Helsen is a doctoral student, The Wharton School, University of Pennsylvania.

The authors extend their appreciation to Wagner Kamakura for supplying the personal computer and mainframe versions of his Segmenter program. They also appreciate the useful comments of both Wagner Kamakura and Michael Hagerty on a previous draft of the article.

More recently, two innovative approaches to traditional full-profile conjoint analysis have appeared in the literature. First, Hagerty (1985) has suggested the use of Q-type factor analysis as a way of finding optimal weights for transforming individual responses that are later analyzed by conventional conjoint methods such as OLS regression or LINMAP (Srinivasan and Shocker 1973). He makes the important point that data pooling of similar respondents' conjoint responses often can reduce variance in the subsequently estimated individual-respondent part worths without unduly increasing the bias of the estimates. He shows, by both synthetic data and empirical data analyses, that his factor analytic approach can indeed improve predictive accuracy at the individual-respondent level.

Kamakura (1988) also is concerned with improving predictive validity; his approach is to form respondent clusters that optimize segment-level part worth estimates. His innovative procedure obtains clusters of respondents and part worth estimates in a single analytic step. It is in contrast to a more traditional, two-stage approach in which each individual's part worths are obtained first, then respondents are clustered in terms of similarities in their part worths. Kamakura's method avoids the pooling of unreliable portions of respondents' part worth estimates. He illustrates, with both synthetic and empirical data, the value of his method for improving accuracy in holdout sample predictions.

Both approaches are designed to improve predictive

Journal of Marketing Research
Vol. XXVI (August 1989), 346–50

accuracy when individual respondents' part worths are used to predict responses to a holdout sample. Hagerty does this by transforming each respondent's conjoint response data to reflect commonalities of the respondent with all other respondents in the sample before computing individualized part worths for that respondent. Kamakura does this by grouping respondents by similarities in their conjoint responses. This grouping is done in such a way as to maximize predictive validity when each cluster's averaged part worths are used to predict individual holdout-sample responses for each member of the cluster.

We find both approaches impressive, both conceptually and operationally. The question we address is: In practice do these methods result in major gains in predictive accuracy in comparison with conventional part worth estimation approaches? We include as "conventional" approaches both full-profile conjoint decomposition and the even simpler self-explicated part worth model (Leigh, Mackay, and Summers 1984) which, incidentally, serves as a first-stage input to hybrid conjoint models (Green 1984).

EXPERIMENTAL DESIGN AND DATA COLLECTION

The stimulus context of our experiment consisted of descriptions of privately offered, unfurnished student apartments in the vicinity of a large eastern university. Subjects for the experiment were undergraduate and graduate business students, most of whom were already living in a student apartment or were considering renting one during the next school year.

The databank consisted of (1) self-explicated attribute-level acceptability ratings and attribute importance values, (2) likelihood-of-renting ratings on 18 calibration profiles (condition 1 or 2), and (3) likelihood of renting on a common set of 16 validation profiles.

Design of the Calibration Set

Table 1 shows the attributes and levels used in the conjoint designs. An orthogonal main effects design of 18 combinations was selected for experimental condition 1, representing a typical calibration set (see Table 2), used to obtain the initial part worths. Condition 2 (see Table 2) was identical to condition 1 with the exception of price. Price was modified in condition 2 to reflect the "quality" of the nonprice part of each profile, subject to maintaining the same range of price and an equal number (six) of cases for each price level. In condition 2, high quality apartments tend to cost more to rent. Subjects were assigned randomly to the profile sets of either condition 1 or condition 2.

Table 1
ATTRIBUTES AND LEVELS USED IN CALIBRATION PHASE OF CONJOINT STUDY

A. Walking time to classes	D. Condition of apartment
1. 10 minutes	1. Newly renovated throughout
2. 20 minutes	2. Renovated kitchen only
3. 30 minutes	3. Poor condition
B. Noise level of apartment house	E. Size of living/dining area
1. Very quiet	1. 24 by 30 feet
2. Average noise level	2. 15 by 20 feet
3. Extremely noisy	3. 9 by 12 feet
C. Safety of apartment location	F. Monthly rent (utilities included)
1. Very safe location	1. $540
2. Average safety	2. $360
3. Very unsafe location	3. $225

Table 2
PROFILE DESCRIPTIONS FOR CALIBRATION CONDITIONS 1 AND 2
(see Table 1 for attribute/level descriptions)

Profile	*A*	*B*	*C*	*D*	*E*	*F-I*	*F-II*[a]
1	1	1	1	1	1	1	1
2	2	2	3	2	2	2	3
3	3	3	2	3	3	3	3
4	1	2	2	2	3	1	2
5	2	3	1	3	1	2	2
6	3	1	3	1	2	3	2
7	1	3	3	2	1	3	2
8	2	1	2	3	2	1	3
9	3	2	1	1	3	2	2
10	1	3	2	1	2	2	1
11	2	1	1	2	3	3	1
12	3	2	3	3	1	1	3
13	1	1	3	3	3	2	3
14	2	2	2	1	1	3	1
15	3	3	1	2	2	1	2
16	1	2	1	3	2	3	1
17	2	3	3	1	3	1	3
18	3	1	2	2	1	2	1

[a]Stimulus descriptions are identical across profiles 1, 3, 5, and 9.

Respondent Tasks

In phase 1 all respondents first responded to the self-explicated tasks, the first step used in hybrid conjoint analysis. For each attribute of Table 1 the respondent was asked to rate its acceptability on a 0–10, equal-interval rating scale. Each respondent then was asked to allocate 100 points across the six attributes so as to reflect their relative importance (constant sum scale).

In phase 2, each respondent received (in randomized order) 18 calibration full-profile cards made up according to the design of Table 2. (As noted before, respondents were assigned randomly to calibration condition 1 or 2.) In each case the respondent was asked to indicate the likelihood (on a 0–100 scale) of renting an apartment of that description, assuming he or she was in need of an apartment close to the university.

In phase 3, each respondent received (in randomized order) 16 validation profiles, each made up of levels 1 and 3 of the attributes in Table 1. Each subset of eight profiles represented a separate orthogonal design, based on a one-eighth replicate of a 2^6 factorial design. (Price

varied independently in half of the profiles and covaried with nonprice levels in the other half.) Each respondent rated each profile in the validation set on the same 0–100 scale that was used in phase 2.

MODEL FITTING

A sample of complete responses was obtained for 99 respondents, 45 of whom received the condition 1 profile cards and 54 of whom received the condition 2 profile cards.

Conjoint Model Fitting

We fit each self-explicated model individually, using the first phase of the hybrid model. Acceptability ratings were normalized to vary between zero and one, as were the importance values; in addition, the importance values were normalized to sum to unity. We first consider a stimulus profile h made up of levels given by the vector

$$\mathbf{i}^{(h)} \equiv [i_1, i_2, \ldots, i_j, \ldots, i_j]$$

in which $i_j^{(h)}$ denotes level $i_j (i_j = 1, I_j)$ of attribute $j (j = 1,J)$, appearing in profile h. We let

$u_{ijk} \equiv$ respondent k's $(k = 1,K)$ normalized self-explicated acceptability score for level i of attribute j and

$w_{jk} \equiv$ respondent k's normalized self-explicated importance weight for attribute j.

We then can find the respondent's self-explicated utility for profile $\mathbf{i}^{(h)}$ from

$$U_{i_1 i_2 \ldots i_J k} = \sum_{j=1}^{J} w_{jk} u_{i_j k}.$$

Hence, a self-explicated prediction of respondent k can be obtained for any profile in the validation set.

Fitting of the conjoint calibration model (condition 1 or 2) was carried out, in the usual way, by OLS regression, in which each profile of Table 2's design was expressed as a set of dummy variables. A main-effects-only model was fit for each respondent. Each respondent's model then was used to predict each of the 16 validation profiles (that were common across all respondents).

Fitting the Hagerty Optimal Weighting Model

Hagerty describes his approach in detail (1985, p. 170–3) and it need not be repeated here. Suffice it to say that Hagerty applies principal components analysis to a matrix of correlations among respondents' response data, in this case their (standardized) vector of responses to the 18 calibration stimuli of either condition 1 or condition 2. This approach is akin to taking the largest R components in a singular value decomposition of the matrix of subjects' standardized (within-subject) responses.

Hagerty uses Cattel's scree test (a graphic procedure) to select the appropriate number of principal components. He mentions, however, that his approach is not sensitive to the number of components extracted. Hagerty suggests that, after the transformed response vector is found for each respondent, the usual methods (e.g., OLS dummy variable regression) can be used to find the part worths. However, he makes the point (p. 180) that optimal weighting will be least useful when (1) response error is unusually high, (2) the design matrix is highly collinear, or (3) respondents are very heterogeneous.

In the present application the design matrix was orthogonal in condition 1 and "almost" orthogonal in condition 2 (correlations due to price averaged out at only −.08). The principal component solution in condition 1 resulted in nine eigenvalues exceeding 1.0 and accounting for 92.5% of the variability in the input matrix. In condition 2 there were 11 such eigenvalues exceeding 1.0 and accounting for 93.3% of the variability in the input matrix.

To test Hagerty's conjecture that his approach would not be sensitive to the selected rank approximation, we set up two sets of predictions for each condition. In condition 1, part worths were computed for rank 4 (low) and rank 9 (high) approximations; in condition 2, part worths were computed for rank 5 (low) and rank 11 (high) approximations.

Fitting the Kamakura Optimal Clustering Model

Kamakura's approach is described in detail in his article (1988, p. 158–62) and is not repeated here. (Professor Kamakura graciously supplied his computer program for implementing his model.) Kamakura's objective is to improve predictive validity by clustering respondents into homogeneous groups and then computing part worth estimates for data pooled within cluster. As he explains (p. 166):

> Rather than grouping consumers on the basis of their similarity in terms of part worth *estimates* (i.e., ignoring the estimation errors), we form segments that maximize the predictive validity of the segment-level part worth estimates—that is, they are adequate representations of the true underlying utility function of each member of the sample.

Kamakura uses an agglomerative, hierarchical grouping procedure in accomplishing this objective.[1]

As a "control" case, against which we could compare the effectiveness of the Kamakura approach, we simply clustered the *original* responses (after standardization within respondent) to the calibration task of 18 evaluations, by condition. Ward's (1963) minimum variance agglomerative routine was used to obtain clusters for the control case. We cut the hierarchical trees at the same number of clusters as were obtained from Kamakura's computer program. To test the sensitivity of the results

[1]Before Kamakura's algorithm was applied, each respondent's data were standardized to zero mean and unit standard deviation (similar to what was done for the Hagerty approach).

to the number of segments, we cut both sets of trees at three and five clusters.

STUDY RESULTS

Three measures of cross-validation accuracy were employed: (1) product moment correlation, (2) mean squared error, and (3) first-choice hits between actual and predicted responses for the 16 validation profiles.[2] All measures were computed at the individual level, before summarization. Analyses were performed by condition number, 1 or 2.

Given the correlated nature of the models (e.g., each set of predictions was compared with the same dependent variable), we used repeated measures ANOVA and Friedman's nonparametric ANOVA to analyze the product moment and mean-squared-error measures. Cochran's nonparametric test for K related samples was used to analyze the first-choice hit data.

We describe the results by response measure (product moment correlation, mean squared error, and first-choice hits). Table 3 is a full summary of the results.

Product Moment Correlation

As noted from Table 3, in condition 1 (based on a completely orthogonal design), no significant differences are found across the eight "models." In particular, the "standard" conjoint model does as well (or better on a sample basis) as either the Hagerty or the Kamakura model. In condition 2 the eight models show correlations within the range of .723 to .744. Clearly, there is no practical difference and there is no statistically significant difference, either. We conclude that all eight approaches lead essentially to the same cross-validation results in terms of the correlation coefficient.

[2]In actuality the 16 validation samples were analyzed by first half versus second half so that the "first-choice hits" pertain to getting one of eight possibilities correct; that is, the expected percentage (under chance) is 12.5%.

Mean Squared Error

In condition 1 we also note from Table 3 that there is no significant difference across models with respect to mean squared error. On a sample basis, OLS regression has the lowest MSE in condition 1.

In condition 2, we again note that no significant differences are found. On a sample basis the self-explication model has the smallest MSE.

First-Choice Hits

In condition 1 we note from Table 3 that Cochran's test does not lead to significant differences among predictive models. The sample-based range is 53.3 to 65.6%. In condition 2 (where no significant differences are noted), the range of first-choice hits is 54.7 to 63.9%. Overall, no model, on a sample basis, consistently outperforms the rest across the six combinations of response measure and experimental condition.

CONCLUSIONS

As shown in Table 3, the traditional OLS regression approach (and the self-explicated models as well) perform at about the same levels as the Hagerty and Kamakura approaches. Ward's clustering program also does about as well as the Kamakura clustering algorithm.

Why is this so? Clearly, we have analyzed only one dataset, based only on a relatively small sample (N = 99). Even so, the sample size is comparable to the sample sizes employed in Hagerty's empirical analysis (N = 110) and Kamakura's empirical study (N = 105). In fairness to the authors, however, neither Hagerty nor Kamakura suggests that dramatic improvements should be associated with his approach.

Hagerty's study (job characteristics of prospective employers) shows significant gains for optimal weighting over OLS individualized regression, but his holdout sample consists of only two profiles (versus the 16 profiles used here). Kamakura's study (checking account attributes) employs a holdout sample of nine profiles; he finds a significant difference (alpha = .10) favoring his

Table 3

DESCRIPTIVE SUMMARY OF MODEL FITTING BY CONDITION AND RESPONSE MEASURES

	Condition 1 (N = 45)[a]			*Condition 2 (N = 54)*[a]		
Model	*Product moment corr.*	*MSE*	*First-choice hits*	*Product moment corr.*	*MSE*	*First-choice hits*
OLS regression	.750[b]	504.3	62.3%	.723	585.4	54.7%
Self-explicated	.745	625.9	62.2	.732	511.2	63.9
Hagerty—high rank	.735	589.2	56.7	.725	772.3	58.3
Hagerty—low rank	.746	600.8	62.2	.731	811.9	57.4
Ward—5 cluster	.706	591.9	53.3	.744[b]	516.2	62.9
Ward—3 cluster	.705	590.8	55.6	.741	537.6	58.3
Kamakura—5 cluster	.716	572.9	65.6	.737	565.4	63.0
Kamakura—3 cluster	.704	590.0	58.9	.729	585.2	58.3

[a]No significant differences are noted across models for each of the three response measures (alpha = .05).
[b]Underline denotes best performance on a sample basis.

model over the procedure that clusters on computed part worths. As he mentions, however, the differences in predictive validity are not striking in practical terms.[3]

Hagerty sums up the matter of conceptual versus practical impact nicely (p. 182) by recommending that his model be used as a preliminary procedure before analysis of the data via either OLS regression or other methods (e.g., constrained regression).[4] Though we consider only one dataset here, our results tentatively indicate the robustness of conventional OLS regression and, to some extent, the self-explicated model as well. Clearly, additional empirical comparisons are needed.

Our empirical findings in no way diminish the originality and creativity of the Hagerty and Kamakura contributions. Indeed, we were motivated to conduct the experiment primarily because of the intellectual interest stimulated by their proposals. Our results simply serve to caution researchers who might otherwise overestimate potential gains in cross validity that might be ascribed to the more sophisticated Hagerty or Kamakura procedure. Further study is needed of the conditions under which these approaches can lead to substantially higher predictive validity than can be obtained with conventional models.

REFERENCES

Akaah, Ishmael P. and Pradeep K. Korgaonkar (1983), "An Empirical Comparison of Predictive Validity for Self-Explicated, Huber-Hybrid, Traditional Conjoint, and Hybrid Conjoint Models," *Journal of Marketing Research*, 20 (May), 187–97.

Cattin, Philippe, Alan Gelfand, and Jeffrey Danes (1983), "A Simple Bayesian Procedure for Estimation in a Conjoint Model," *Journal of Marketing Research*, 20 (February), 29–35.

Currim, Imran S. (1981), "Using Segmentation Approaches for Better Prediction and Understanding from Consumer Mode Choice Models," *Journal of Marketing Research*, 18 (August), 301–9.

Green, Paul E. (1977), "A New Approach to Market Segmentation," *Business Horizons*, 20 (February), 61–73.

——— (1984), "Hybrid Models for Conjoint Analysis: An Expository Review," *Journal of Marketing Research*, 21 (May), 155–9.

——— and Wayne S. DeSarbo (1979), "Componential Segmentation in the Analysis of Consumer Trade-Offs," *Journal of Marketing*, 43 (Fall), 83–91.

———, Stephen M. Goldberg, and Mila Montemayor (1981), "A Hybrid Utility Estimation Model for Conjoint Analysis," *Journal of Marketing*, 45 (Winter), 33–41.

——— and Abba M. Krieger (1989), "A Hybrid Conjoint Model for Price-Demand Estimation," *European Journal of Operational Research*, 37, in press.

——— and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research*, 5 (September), 103–23.

Hagerty, Michael R. (1985), "Improving the Predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis," *Journal of Marketing Research*, 22 (May), 168–84.

Kamakura, Wagner A. (1988), "A Least Squares Procedure for Benefit Segmentation with Conjoint Experiments," *Journal of Marketing Research*, 25 (May), 157–67.

Leigh, Thomas W., David B. Mackay, and John O. Summers (1984), "Reliability and Validity of Conjoint Analysis and Self-Explicated Weights: A Comparison," *Journal of Marketing Research*, 21 (November), 456–62.

Madansky, Albert (1980), "On Conjoint Analysis and Quantal Choice Models," *Journal of Business*, 53 (July), 537–44.

McFadden, Daniel (1980), "Econometric Models for Probabilistic Choice Among Products," *Journal of Business*, 53 (July), 513–29.

Moore, William (1980), "Levels of Aggregation in Conjoint Analysis: An Empirical Comparison," *Journal of Marketing Research*, 17 (November), 516–23.

Srinivasan, V. (1980), "Comments on Conjoint Analysis and Quantal Choice Models," *Journal of Business*, 53 (July), 547–50.

——— and Allan D. Shocker (1973), "Linear Programming Techniques for Multidimensional Analysis of Preferences," *Psychometrika*, 38, 337–69.

Ward, Joe H. (1963), "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, 58, 236–44.

[3]We reiterate that Kamakura's control case entails a two-step procedure in which clusters are formed across individuals' already-computed individual part worths; cluster averages then are used to obtain predictions. Our control case is more in keeping with Kamakura's desire to cluster on the original responses (not the derived part worths).

[4]In this study the attributes tended to be "monotonic," for example, higher safety preferred to lower safety. Therefore the homogeneity across respondents tends to be higher than would otherwise be the case. Higher homogeneity favors the Hagerty model; hence, our findings are conservative.

Reprint No. JMR263107