Conjoint Internal Validity Under Alternative Profile Presentations

PAUL E. GREEN
KRISTIAAN HELSEN
BRUCE SHANDLER*

Recent research suggests that using orthogonal arrays in full-profile conjoint may fail to provide adequate internal validity when the validation set consists of Pareto optimal profiles or when prices covary with the rest of the profiles' attribute levels. This study's findings indicate that partworths calibrated in the wrong environment predict a holdout sample as well as those calibrated in the correct (validation) environment do.

Since the introduction of conjoint analysis to marketing research in the early 1970s (Green and Rao 1971; Johnson 1974), academic and industry researchers have continually called for more research on conjoint reliability and validity. Researchers are exploring a variety of questions ranging from attribute order bias to partworth comparisons across geographic areas (e.g., Chapman and Bolton 1985; Huber and Hansen 1986; Levin et al. 1983; Umesh 1987).

This study deals primarily with the question of how well orthogonal arrays with independently varying price levels predict holdout profiles in which price covaries with the rest of the profile and all nonprice attributes are Pareto optimal (i.e., no profile is dominated by another with respect to attribute-by-attribute partworths).

INDEPENDENT VERSUS COVARYING PRICE

In virtually all conjoint applications reported in the literature, orthogonal arrays are used to implement full profile presentations (Green and Srinivasan 1978). If price is one of the attributes, its levels are typically varied independently of the other attributes. The same is true if the trade-off matrix approach is used.[1] An implication of this independent variation is that some of the conjoint profiles could look unreasonable; for example, one could see high quality profiles at bargain basement prices or low quality profiles priced too high. Repondents might be prone to distort their preference evaluations in trying to react to the profiles' inconsistencies.

PARETO OPTIMALITY

In some conjoint studies, many (or all) of the attributes may be monotonic in the sense that most respondents would order the attribute levels the same way; that is, "more is better." Examples include product durability, length of guarantee, price, convenience of application, and so on. If such is the case, some profiles could dominate others in an orthogonal array (i.e., some profiles may be at least as good as the dominated profile on all attributes and strictly better on at least one attribute). Although this situation would make the preference task easier to perform, in the real marketplace, buyers would not be likely to find dominances; and even if dominances were found, buyers would be presumed to recognize them and to eliminate dominated options at the outset.

Several researchers also have explored the tendency for Pareto optimal sets to exhibit negative correlations across attribute sets (Curry and Faulds 1986; Krieger and Green 1987; McClelland 1978; Newman 1977; Stillwell, Seaver, and Edwards 1981). In one of the more recent papers on the topic, Johnson, Meyer, and Ghose (1986) explore some of the issues related

*Paul E. Green is the S.S. Kresge Professor of Marketing and Kristiaan Helsen is a doctoral candidate in Marketing, both at The Wharton School, Philadelphia, PA 19104. Bruce Shandler is Vice President of Intersearch Corporation, a marketing consulting firm located at 440 Lombard St., Philadelphia, PA 19147.

[1]Several researchers have discussed problems associated with correlated attributes, particularly price (e.g., Goldberg, Green, and Wind 1984, p. S113; Green and Srinivasan 1978, p. 110; Louviere and Hensher 1983, p. 352; and Mahajan, Green, and Goldberg 1982, p. 335). Nonetheless, not all conjoint tasks are amenable to the authors' solutions, and various software packages (e.g., Sawtooth Software's Adaptive Conjoint Analysis) continue to treat price as an independently varying attribute.

© JOURNAL OF CONSUMER RESEARCH • Vol. 15 • December 1988

to compensatory models (such as those used in conjoint analysis and logit/probit choice modeling). They show, via computer simulation, that a compensatory model calibrated in an orthogonal environment is not a good predictor of the true compensatory model as parameterized in a Pareto optimal environment.

On the face of it, the Pareto optimal and correlated price issues would seem to pose serious threats to the predictive validity of orthogonal main effects plans with independently varying price. The main objective of this study is to assess how serious these issues are to internal predictive validity.[2]

PREVIOUS RESEARCH

Johnson et al.'s (1986) research is particularly germane to the problems addressed in this article. Their research suggests the plausible argument that a model should be calibrated on stimuli similar to those that would be encountered in the real world. One would think that real-world options would already be Pareto optimal and would also reflect covarying price, i.e., high quality products would simply cost more.

However, Pareto optimality across nonprice attributes and covarying price make the profiles increasingly difficult to judge; the range in utilities across options becomes much more narrow. The resulting preference judgments may be more prone to error. Huber and Hansen (1986) explored this issue in the context of Sawtooth's Adaptive Conjoint Analysis package (which is based on graded paired comparisons). They found that the predictive accuracy of profile pairs that were balanced to exhibit small differences in utility actually produced higher internal validity (even though the choices were more difficult to make).

Huber and Hansen chose a validation set that consisted of close pairs with respect to utility.[3] Two concluding points of Huber and Hansen's study are of particular interest (1986, p. 163):

> (1) Thus, having trouble in deciding between alternatives appears to lead to greater richness in response and greater correspondence with subsequent choice, rather than the reverse.
>
> (2) It is reasonable to expect that more complex holdout choices would be best predicted by conjoint tasks that elicit analogously complex processing.

EXHIBIT

ATTRIBUTES AND LEVELS USED IN CONJOINT STUDY

A. Walking time to classes
   1. 10 minutes
   2. 15 minutes
   3. 20 minutes
   4. 30 minutes

B. Noise level of apartment house
   1. Very quiet
   2. Average
   3. Noisier than average
   4. Very noisy

C. Safety of apartment location
   1. Very safe
   2. Average
   3. Less safe than average
   4. Very unsafe

D. Cleanliness of apartment building
   1. Very clean
   2. Average
   3. Worse than average
   4. Very dirty/messy

E. Condition of apartment
   1. Newly renovated throughout
   2. Renovated kitchen only
   3. Fair condition
   4. Poor condition

F. Size of living/dining area
   1. 24 by 30 feet
   2. 15 by 24 feet
   3. 12 by 15 feet
   4. 9 by 12 feet

G. Monthly rent (utilities included)
   1. $225
   2. $270
   3. $315
   4. $360
   5. $405
   6. $450
   7. $495
   8. $540

If replicated in other contexts, their findings (and those of Johnson et al. 1986) would suggest two practical implications for full profile conjoint: where possible, orthogonal arrays should be modified so that the profiles are Pareto optimal with respect to nonprice attributes, and price should covary with the quality of the rest of the profile.

DESIGN OF THE STUDY

The stimulus context for the study consisted of descriptions of privately offered, unfurnished, student apartments located near a large, Eastern university. Subjects for the experiment were undergraduate and graduate business students, most of whom were either already living in a student apartment or were contemplating renting one during the next school year.

Stimulus Design

The Exhibit shows the attributes and levels used in the basic conjoint design. An orthogonal main effects

[2]Even more importantly, of course, is how serious each of the conditions is for external validity. This article deals with internal validity on the assumption that if orthogonally designed calibration models fail here, one might expect even poorer performance in predicting marketplace behavior.

[3]It should also be noted that in Huber and Hansen's study some stimulus pairs were manipulated to share some of the same attribute levels, making such paired comparisons easier to evaluate (i.e., some of the pairs were similar perceptually and similar in utility).

plan of 32 profiles was first constructed from the attribute levels; in keeping with conventional conjoint studies, rental price was treated as an independently varying attribute. This array will be called the base case array. The six nonprice attributes and levels (walking time to classes, noise level, safety, cleanliness of building, condition of apartment, size of living/dining area) were developed from a preliminary questioning of students regarding their judgments about important attributes in student apartments. Another consideration in the design was that the ordering of all levels be monotonic (with utility) in each attribute. Student apartment descriptions have served successfully as stimuli in other experiments (e.g., Huber and Hansen 1986; Johnson and Meyer 1984).

In real situations, apartment rental price would be expected to covary with enhancements in each of the nonprice attributes. Information was obtained on associations between features and price from city newspaper and school newspaper classified advertising; in addition, a convenience sample of students was informally questioned on "willingness to pay" for various enhancements to the base levels of each attribute. From these (somewhat ad hoc) sources of information, a cost for each enhancement was computed for each attribute so that the sum of the costs (plus a base rental price level) covered the known range of apartment prices. In this manner, a second array, called the covarying price array, was constructed from the initial array. Each profile was the same as that in the original array except that total apartment rental price covaried with the attractiveness of the nonprice levels. (In keeping with real-world situations, information was not provided on how the total rental price was computed or what the separate attribute-level contributions were.)

The third array, the covarying price/Pareto array, was constructed from the original base case design in two stages. First, the orthogonal design contained 11 dominated profiles in the subset of six nonprice attributes; these dominated profiles were replaced by nondominated profiles, so that the full set of 32 profiles was Pareto optimal with respect to the nonprice attributes. Following this, a rental price was attached to each profile, employing the same assignment procedure used in the covarying price array.

Next, a set of 16 new profiles, called the validation array, was constructed to respect the covarying price/Pareto condition. None of these profiles was the same as any of the profiles used in the three experimental conditions (base case, covarying price, or covarying price/Pareto).

Data Collection

The subjects for the experiment consisted of 120 undergraduate and graduate business students, assigned randomly to each of the three experimental conditions (with 40 subjects assigned to each). All data were collected by self-administered questionnaire. On average, the interview required about 35 minutes of class time to complete. All subjects who completed the interview were eligible to participate in a lottery involving cash prizes totaling $150. Subjects provided evaluated data in two phases.

In Phase I, subjects received 32 full-profile stimulus descriptions drawn from the experimental condition to which they were assigned randomly. Respondents rated each of the 32 calibration profiles on a 0–100 point likelihood-of-consideration scale, assuming that they would be in the market for an apartment during the next semester.

In Phase II, subjects received four sets of four new apartment profiles drawn from the common validation array. Each of the options in each set of four was Pareto optimal with a rental price that covaried with the rest of the profile. Each subject was asked to rank (strictly) each of four options in each validation set in terms of the option's consideration as a rental. Following the ranking, each option was rated on the same 0–100 point likelihood-of-consideration scale used in Phase I. The range of apartment prices per month in each of the four arrays was:

Base Case	$225 to $540
Covarying Price	$225 to $450
Covarying Price/Pareto	$360 to $450
Validation	$360 to $450

RESULTS

Phase I Data Fitting

The Phase I full-profile evaluations of each subject for the 32 conjoint profiles were analyzed by respondent for each of the 120 subjects, using main-effects, dummy-variable regression. As noted earlier, the covarying price and covarying price/Pareto conditions have no dummy variables for price because of the dependence of price on the rest of the profile. Thus, these two conditions have seven additional degrees of freedom. Since the number of degrees of freedom for error in the regressions varied across experimental conditions (*df* = 6 for the base case and *df* = 13 for the covarying price and covarying price/Pareto), adjusted R^2s were computed as appropriate measures of fit. The following tabulation shows the results ($n = 40$).

	Base case	*Covarying price*	*Covarying price/Pareto*
Adjusted R^2			
Mean (across subjects)	0.713	0.454	0.439

The base case (adjusted) R^2 is 0.713, versus only 0.454 and 0.439 for covarying price and covarying price/Pareto, respectively. As shown in the following tabulation of the ANOVA summary, the result is significant at the 0.001 level.[4]

	df	*SS*	*MS*	*F*	*Probability*
Conditions	2	1.909	0.954	26.82	0.001
Error	117	4.164	0.036		

Not surprisingly, we found that the average coefficient of variation in the dependent variable was significantly higher under the base case condition ($p < 0.01$), suggesting that subjects were using a greater range on the 0–100 likelihood rating scale.[5]

Phase II Validation

As described earlier, in Phase II each respondent received the same set of 16 validation profiles, constructed to respect the covarying price/Pareto condition. Each respondent saw four sets of four profiles each; options were ranked within set and then individually rated on a 0–100 likelihood-of-consideration scale. Four internal validation measures were selected: the product moment correlation, the RMSE measure, the Spearman rank correlation, and the number of first-choice hits. The Table summarizes the results.[6]

As noted in the Table, we found no significant differences across experimental conditions for each of the four internal validation measures. Apparently, the base case approach predicts holdout samples as well as the approach in which price covaries and the nonprice attributes are all Pareto optimal.

The RMSE measure provides a good index of error remaining in the dependent variable after the predictions. The RMSEs are about the same across all three experimental conditions. Although the sample product moment correlations and Spearman rank correlations are higher for the base case condition, this finding is not statistically significant.

TABLE

RESULTS OF PHASE I PREDICTIONS OF PHASE II RESPONSES

Internal validation measure	Condition: Base case	Covarying price	Covarying price/Pareto	*p*
Product moment correlation				
Mean (across subjects)	0.558	0.425	0.428	*ns*
Root-mean-squared error				
Mean (across subjects)	21.47	23.34	21.30	*ns*
Spearman rank correlation				
Mean (across blocks and subjects)	0.532	0.423	0.405	*ns*
Number of first-choice hits (maximum is 4)				
Mean (across subjects)	1.95	1.75	1.73	*ns*

Note: n = 40 per condition and *ns* denotes not significant at the 0.05 level in a three level, one factor ANOVA.

DISCUSSION

In contrast to the findings of Huber and Hansen (1986), the results of this study provide no empirical support for higher internal validity when the model used to make up the validation set of stimuli is also used to construct the calibration stimuli. We speculate that in Phase I, the base case generally entails easier evaluations in which utility differences tend to be larger, leading to higher adjusted R^2s. The dependent variable range is less restricted here than in the varying price conditions. Moreover, the range in the price variable is greater in the orthogonal case than in the covarying price/Pareto condition; hence, in this latter case, we also have a range restriction on one of the predictor variables. What about the Phase II predictions? In Phase II, all respondents received the same 16 validation profiles. However, the predicted values in the covarying price/Pareto condition still reflect range restrictions as related to the calibration profiles.

Implications for Pareto Optimality Designs

What are the implications of this experiment for modifying traditional orthogonal designs to incorporate Pareto optimality? On the one hand, the present experiment suggests that Pareto optimal designs do not lead to significantly better predictions of holdout profiles whose nonprice attributes are Pareto optimal. On the other hand, we have found elsewhere that in relatively large orthogonal designs, finding arrays that are both orthogonal and Pareto optimal is not too difficult. In general, as the number of attributes and levels within attributes increase, it becomes easier to convert an orthogonal array into one that is orthogonal and Pareto optimal.

We suggest that Pareto optimality is a rather weak property that often can be achieved in applications of practical interest without sacrificing orthogonality. (In other cases, it can be achieved with designs that are almost orthogonal.) In view of the ease with which

[4]Similar conclusions were reached following a preliminary transformation of the correlations via Fisher's Z transformation. (Throughout, all ANOVAs were replicated with Fisher's Z without change in the substantive conclusions.)

[5]A three level, one factor ANOVA was also carried out in which the response variable was the RMSE (root-mean-squared error) for each subject. The results of this ANOVA showed no significant differences in average RMSE across conditions ($p > 0.05$), supporting the conjecture that the ANOVA results reflect higher dependent variable variance in the base case condition.

[6]Although not shown here, the response data of Phase II were also tested for response block effect (i.e., to see if average rankings differed across the four blocks). No significant block effects were found at the 0.05 alpha level.

computer search heuristics can produce designs that are orthogonal and Pareto optimal, we think it is a good idea to seek designs that fulfill this objective. Unlike the varying price condition, Pareto optimality need not lead to designs where the alternatives are highly negatively correlated. (However, see Wiley 1977 for a method of constructing Pareto optimal designs that does tend to produce negative correlations among the attributes.)

Implications for Embedding Price in the Conjoint Profile

The varying price condition appears to raise more serious research questions. First, allowing price to covary with the rest of the profile effectively embeds price in the partworths of the nonprice attributes. Perfect multicollinearity results if price is functionally related to the rest of the profile; price-level parameters are not estimated separately.

In some applied problems, the researcher may wish to add an orthogonal component to the embedded price to examine the utility ascribed to departures from base levels (see Goldberg et al. 1984, p. S113). Overall price would still be correlated, but not perfectly so. However, we have found that, in many cases, subjects' evaluations are not sensitive to the orthogonally designed deviations around the base levels that are, themselves, varying over a much wider range.

Of further interest is the fact that covarying price may take two forms. In the present example, the respondent received no information regarding how each enhancement of the nonprice attributes contributed to increased price. Unfortunately, many real pricing situations are such that the consumer does not know how each component of the product influences price. In other applications, the seller may offer a product that does provide information on each component's contribution to the overall price. Examples include personal computers for which buyers see the components (e.g., color monitor, printer, expansion slots, hard disk capacity) separately priced but may purchase the computer "package" at a discount from the sum of the component prices.

CONCLUDING CAVEATS

In contrast to the previous research by Johnson et al. (1986) and Huber and Hansen (1986), the present study does not find response differences between orthogonal and covarying price/Pareto designed stimuli.[7] Clearly, additional experiments are needed to gain information about the conditions that do produce higher validation for covarying price/Pareto conditions or, in general, for stimulus profiles that are closer in utility (and, hence, more difficult to evaluate).

As Huber (1987) has suggested, the use of orthogonal designs may provide a high degree of robustness over various task simplifications (e.g., ignoring levels and/or entire attributes) that subjects may employ in coping with the job of profile evaluation. In particular, orthogonal designs could guard against possible sources of misspecification error that may occur when various simplifying decision strategies are employed.

The research reported here should be considered as preliminary. Several caveats underlie the present study's findings.

1. Three samples of only 40 subjects each were tested. Given the type of cross-subject design employed here, future studies would benefit not only from larger samples, but also from extensions to nonstudent populations as well.

2. As noted, the predictions in this study were considerably below those typically found in other conjoint studies. The respondent tasks were extensive (entailing 32 calibration profiles and 16 validation profiles). Simpler tasks could be designed for future investigations.

3. The study probably would have benefited from the collection of other kinds of reaction measures (Huber and Hansen 1986), such as measures of subject interest in the task, stimulus believability, decision difficulty, and so on.

4. The use of an orthogonal main effects design (i.e., the base case) could be unduly restrictive. Although this type of design is often used in applied conjoint studies, other experimental designs that either protect main effects against two-way interaction confounding or actually estimate selected two-way interactions (Carmone and Green 1981; Louviere 1988) should be explored.

Meanwhile, the achievement of Pareto optimality and orthogonality often is possible in larger experimental designs (e.g., designs with nine to 10 attributes, with each attribute having three to four levels). The question of covarying price is a more complex issue, particularly when the respondent is not told how overall price relates to the nonprice attributes that make up the rest of the profile. It is not clear from the empirical evidence assembled to date that covarying price invariably leads to higher internal validity

[7]In the case of the Johnson et al. (1986) report, many of their results were based on simulations of synthetic data; in contrast, our study attempts to deal with empirical data collection with all of the attendant problems of experimental control and potential sources of noise.

and better understanding of substantive findings.[8] Our knowledge about the influence of covarying price/Pareto optimality on external validation is even more inchoate.

[*Received October 1987. Revised May 1988.*]

REFERENCES

Carmone, Frank J. and Paul E. Green (1981), "Model Misspecification in Multiattribute Parameter Estimation," *Journal of Marketing Research,* 18 (February), 87–93.

Chapman, Randall G. and Ruth N. Bolton (1985), "Attribute Presentation Order Bias and Nonstationarity in Full Profile Conjoint Analysis Tasks," in *AMA Educators Conference Proceedings,* eds. Robert F. Lusch et al., Chicago, IL: American Marketing Association, 373–379.

Curry, David J. and David J. Faulds (1986), "Indexing Product Quality: Issues, Theory, and Results," *Journal of Consumer Research,* 13 (June), 134–145.

Goldberg, Stephen M., Paul E. Green, and Yoram Wind (1984), "Conjoint Analysis of Price Premiums for Hotel Amenities," *Journal of Business,* 57 (January), S111–S132.

Green, Paul E. and Vithala R. Rao (1971), "Conjoint Measurement for Quantifying Judgmental Data," *Journal of Marketing Research,* 8 (August), 355–363.

——— and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research,* 5 (September), 103–123.

Huber, Joel (1987), "Conjoint Analysis: How We Got Here and Where We Are," in *Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis and Computer Interviewing,* ed. Richard M. Johnson, Ketchum, ID: Sawtooth Software, 237–252.

——— and David Hansen (1986), "Testing the Impact of Dimensional Complexity and Affective Differences of Paired Concepts in Adaptive Conjoint Analysis," in *Advances in Consumer Research,* Vol. 14, eds. Melanie Wallendorf and Paul Anderson, Provo, UT: Association for Consumer Research, 159–163.

Johnson, Eric J. and Robert J. Meyer (1984), "Compensatory Choice Models of Noncompensatory Processes: The Effect of Varying Context," *Journal of Consumer Research,* 11 (June), 528–541.

———, Robert J. Meyer, and Sanjoy Ghose (1986), "When Choice Models Fail: Compensatory Representations in Efficient Sets," working paper, Carnegie-Mellon University, Pittsburgh, PA 15213.

Johnson, Richard M. (1974), "Trade-off Analysis of Consumer Values," *Journal of Marketing Research,* 11 (May), 121–127.

Krieger, Abba M. and Paul E. Green (1987), "Differential Weighting in Multiattribute Choice Models," working paper, Marketing Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104.

→ Levin, Irwin P., Jordan J. Louviere, Albert A. Schepanski, and K.L. Norman (1983), "External Validity Tests of Laboratory Studies of Information Integration," *Organizational Behavior and Human Performance,* 31, 173–193.

Louviere, Jordan J. (1988), *Analyzing Decision Making: Metric Conjoint Analysis,* Beverly Hills, CA: Sage.

——— and David A. Hensher (1983), "Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event," *Journal of Consumer Research,* 10 (December), 348–361.

Mahajan, Vijay, Paul E. Green, and Stephen M. Goldberg (1982), "A Conjoint Model for Measuring Self and Cross Price-Demand Relationships," *Journal of Marketing Research,* 19 (August), 334–342.

McClelland, Gary H. (1978), "Equal versus Differential Weighting for Multiattribute Decisions: There Are No Free Lunches," Report No. 207, Center for Research on Judgment and Policy, University of Colorado, Boulder, CO 80309.

→ Newman, J. Robert (1977), "Differential Weighting in Multiattribute Utility Measurement: Where It Should and Where It Does Make a Difference," *Organizational Behaviour and Human Performances,* 20 (December), 312–325.

→ Stillwell, William G., David A. Seaver, and Ward Edwards (1981), "A Comparison of Weight Approximation Techniques in Multiattribute Utility Decision Making," *Organizational Behavior and Human Performance,* 28 (August), 62–77.

→ Umesh, U.N. (1987), "Transferability of Preference Model Across Segments and Geographic Areas," *Journal of Marketing,* 51 (January), 59–70.

Wiley, James B. (1977), "Selecting Pareto Optimal Subsets from Multiattribute Alternatives," in *Advances in Consumer Research,* Vol. 5, ed. H. Keith Hunt, Ann Arbor, MI: Association for Consumer Research, 171–174.

[8]As one reviewer pointed out, orthogonal designs (including price) allow the researcher to make predictions for the full Cartesian product set of profiles, but the embedded price condition restricts predictions to the Cartesian product set of profiles composed of the nonprice attributes only.