Adaptive Conjoint Analysis: Some Caveats and Suggestions; Comment
Green, Paul E.; Krieger, Abba M.; Agarwal, Manoj K.; Johnson, Richard M.
JMR, Journal of Marketing Research; May 1991; 28, 2; ABI/INFORM Global
pg. 215

Adaptive Conjoint Analysis: Some Caveats and Suggestions

Researchers in conjoint analysis generally agree that one of the most significant developments in the growth and diffusion of that methodology occurred in the mid-1980s with the introduction of commercial conjoint computer packages. These include Bretton-Clark's Conjoint Designer, Conjoint Analyzer, SIMGRAF, BRIDGER, and Conjoint LINMAP (Herman 1988) and Scott Smith's assorted programs for conjoint analysis and related techniques (Smith 1988).

One microcomputer package that has achieved considerable publicity and commercial application is Richard Johnson's Adaptive Conjoint Analysis (Johnson 1987a). The package is novel in that conjoint data are collected via a computer-based interactive technique that customizes the stimulus presentations that respondents evaluate. A review of Adaptive Conjoint Analysis (ACA) by Carmone (1987) was highly favorable, as have been comments by both academicians (Huber and Hansen 1986) and industry users (Finkbeiner and Platz 1986) of the ACA package.

There is little question about the high popularity of ACA in industry applications. Sawtooth's recent directory of marketing research firms using ACA lists 52 of the leading research agencies in the U.S. More than 200 practitioners have attended each of Sawtooth's annual conferences over the past three years. However, with few exceptions (Green, Krieger, and Bansal 1988; Finkbeiner 1988; Klein 1986), the assumptions underlying ACA have remained largely unquestioned by the research community.

The purpose of our research note is to examine some of these assumptions. This objective is implemented in two ways. First, we briefly review the data collection and measurement aspects of ACA. These steps then are formalized in terms of an additive, main-effects part-worth model that combines self-explicated data on attribute-level preference rankings and attribute importance ratings with graded paired-comparison preferences for...
partial profiles of product concepts. We describe some of the problems that can ensue as the researcher attempts to combine the compositional and decompositional parts of the model in arriving at "final" partworths.

We test our conjectures on a large (N = 170) database involving business students' preferences for apartment descriptions (Agarwal and Green 1989). Specifically, we examine the difficulties that can arise when respondents fail to retain consistency in their attribute-level preferences and use more extreme values on the graded paired comparisons scale than would be predicted by the ACA model.

Finally, we offer what we hope to be constructive suggestions about how ACA's partworth estimation procedure can be improved in future versions of the software. Our critique is limited to the data collection and partworth estimation phase; no comments are offered on ACA's buyer choice simulator, which is also part of the package.

THE ACA DATA COLLECTION AND PARTWORTH ESTIMATION PROCEDURE

Version 2.0 of Sawtooth's ACA, introduced in early 1988, is the version examined here. Its interviewing module consists of four phases (Johnson 1987a):

1. In Phase I, each respondent ranks his or her preferences for each level of each attribute in turn. An additional feature in this phase enables the respondent to declare that some levels are completely unacceptable.

2. In Phase II, the respondent is presented the best and worst levels (obtained from the Phase I rankings) for each attribute in turn. The respondent then rates the importance of the attribute on a 1 to 4 (equal-interval) rating scale where 4 denotes highest importance.

3. In Phase III, the respondent receives a set of paired partial profiles, with members of each pair shown side by side on the computer screen. In each paired comparison the respondent indicates, on a 9-point equal-interval scale (with 9 denoting strongly prefer right), which of the two profiles is preferred and by how much. (ACA later transforms this scale to range from -4 to +4 with zero denoting the indifference point).

4. In Phase IV, the respondent receives from two to nine profile descriptions that are each composed, at most, of eight attributes. Calibration concepts are chosen by the software so as to progress in preference from highly undesirable to highly desirable. The respondent rates each calibration profile on a 0–100 likelihood-of-purchase scale.

Graded Paired Comparisons

The graded paired-comparison section (Phase III) is the heart of the ACA methodology and gives rise to the modifier "adaptive." The procedure is adaptive in the sense that each paired comparison is constructed so as to take advantage of the information obtained about a respondent's partworths in preceding steps. This information is used to select the specific pair of partial profiles that one receives for evaluation at the current step.

It should be noted, however, that Johnson's search routine chooses pairs partially on the basis of the dependent variable (estimated utility) and not solely on the basis of the design matrix of predictor variables. Of course, were pairs chosen randomly, the difference in the self-explicated, partial-profile utilities (computed from Phase I) could be so large as to go beyond the limits of Phase III's rating scale.

If so, the respondent would be forced to rein in the scale values to agree with the range embodied in the Phase III partworth updating program. Fortunately, the heuristic's search for close-in-utility pairs (subject to satisfying other rules that attempt to achieve an "almost orthogonal" design matrix) tends to preclude reaching these extremes. The problem of scale compatibility between Phases I/II and Phase III appears to involve just the opposite case—respondents who see two almost equal-in-utility partial profiles tend to stretch the Phase III scale to try to accommodate the full range of difference. This problem is exacerbated when more than two attributes describe each concept because of the correspondingly higher chance of finding two essentially equal-in-utility partial concepts.

Huber and Hansen (1986) and Agarwal (1989) found empirically that little was gained by constructing paired concepts with more than two (or possibly three) attributes each. Simpler concepts (in contrast to pairs containing as many as five attributes) could have led to higher predictive accuracy for two reasons: greater simplicity for judging and greater utility differences, as obtained from the selection heuristic. Currently, most academic researchers, at least, use only two attributes (or possibly three) for each partial concept in Phase III.

Computing the Partworths

Johnson uses an OLS regression updating procedure (which he calls "Bayesian") to adjust the tentative partworths of Phases I/II to the respondent's judgments of Phase III. Details of the regression updating model are given by Johnson (1987a). The trick, of course, is to have commensurate scale values between Phases I/II and Phase III. As additional Phase III responses are accumulated (and if they are reliable), the Phases I/II partworths receive less weight in the partworth estimation at the end of Phase III.

---

1Recently, Sawtooth Software announced Version 3.0 of ACA, which appeared on the market in late 1989. We comment on one of its features at the end of this note.

2By "design matrix" we mean values of the predictor variables in the OLS regression procedure used to estimate individuals' partworths (for a detailed description, see Johnson 1987a).

3In contrast, Huber and Hansen (1986) found (for a given number of attributes per concept) that smaller utility differences slightly increased the predictive accuracy of first-choice holdout profiles. They ascribe these results to greater respondent interest in more difficult-to-judge pairs. We believe additional empirical research is needed on this point.
A MODEL OF THE ACA PROCESS

We now discuss ACA in more formal terms because, to the best of our knowledge, a formal statement of the model has not yet appeared in print. Assume that a product profile (i.e., concept) is described on M attributes, where the mth attribute has Jm levels; m = 1, ..., M. For an individual respondent we assume a set of partworths \( \beta_{mj} \), where m is the attribute and j is the level within attribute; m = 1, ..., M; j = 1, ..., Jm. We also assume that a respondent's evaluation of a profile \( J_1, ..., J_M \) is a function of

\[
\sum_{m=1}^{M} \beta_{mj} w_m
\]

To estimate \( \beta_{mj} \), ACA employs two types of input—the Phases I/II data on self-explicated preference rankings and attribute importances and the Phase III data on paired comparisons. We let \( w_m \) denote the subject's importance weight for attribute m; in ACA \( w_m \) is an integer: 1, 2, 3, or 4. We let \( r_{mj} \) denote the reflected preference rank assigned by the subject to level \( j \) of attribute \( m \).

We then have as scale values

\[
y_{mj} = w_m (r_{mj} - 1)/(J_m - 1) - .5].
\]

Equation 1 is a general formulation for calculating ACA's Phases I/II scale values.

For example, assume that attribute \( m \) has four levels, receives a weight \( w_m = 3 \), and the four levels have (reflected) preference ranks \( (r_{m1}, r_{m2}, r_{m3}, r_{m4}) = (2, 1, 4, 3); \) then the respective \( (y_{m1}, y_{m2}, y_{m3}, y_{m4}) = (-.5, -1.5, 1.5, .5) \).

For Phases I/II, the model is simply

\[
y_{mj} = \beta_{mj} + \epsilon_{mj},
\]

where the \( \epsilon_{mj} \) are assumed to be i.i.d. \( N(0, \sigma^2) \)—that is, identically, independently, normally distributed with mean zero and variance \( \sigma^2 \).

Phase III provides the second source of data. These data represent responses to \( P \) graded paired comparisons. We assume that a subset of \( \ell \) attributes, \( s_{1}^{(p)}, ..., s_{\ell}^{(p)} \), are selected for paired comparison \( p \), where \( \ell \) can range from two to five attributes. Two “subprofiles” are shown at a time, say \( i_1, ..., i_{\ell}^{(p)} \) and \( k_1, ..., k_{\ell}^{(p)} \), denoting the levels for attributes \( s_1, ..., s_{\ell} \).

The respondent provides a “score” as an integer between 1 and 9. The score indicates the direction and extent to which one profile (left or right side of the screen) is preferred to the other. In preparing the response data for analysis, the integer 5 is subtracted from the raw score, leading to an adjusted score \( z^{(p)} \) that is integer-valued and ranges from -4 to +4. The ACA model for Phase III is written as

\[
z^{(p)} = \sum_{j=1}^{\ell} \beta_{j}^{(p)} b_j^{(p)} - \beta_{j}^{(p)} b_j^{(p)} + \delta^{(p)}
\]

where \( \delta^{(1)}, ..., \delta^{(p)} \) are i.i.d. \( N(0, \sigma^2) \) and also assumed to be independent of \( \epsilon_{mj} \).

Ultimately, equations 2 and 3 are combined (via OLS regression) to estimate \( \beta_{mj} \) from the self-explicated data of Phases I/II and the graded paired comparisons data of Phase III.

AN EMPIRICAL EXAMINATION OF THE PARTWORTH UPDATING PROCEDURE

In the fall of 1989, Agarwal and Green (1989) designed and implemented an ACA-based study entailing student evaluations of apartment descriptions. Six attributes (transport time to campus, noise level, safety, apartment condition, size, and rent) were used in the study; see Agarwal and Green (1989) for detailed descriptions.

Data were collected from 170 business students attending a large eastern university and were made available to us for further analysis. Given the monotonic nature of all six attributes (e.g., less rent preferred to more rent), the researchers employed the ACA option that assumes a common ranking of the levels within each attribute across each respondent. Also, in view of the Huber and Hansen findings, the paired-comparisons portion (Phase III) contained only two attributes per concept. Each respondent received 15 such pairs. Then respondents received two sets of orthogonal holdout profiles of 18 and 16 full-profile concepts, respectively. In each case, respondents rated each holdout profile on a 0–100 likelihood-of-renting scale.

Version 2.0 of ACA contains an “audit trail” file that summarizes, for each respondent, the full set of responses to each of the data collection phases. In this way, an interested researcher can reconstruct the entire process of partworth updating, step by step. This dataset provided us a way to examine a variety of research questions about the ACA model.

Respondent Evaluation of Phase III Pairs

Given our preceding discussion of ACA and the heuristic it employs to obtain Phase III pairs that are as similar as possible in utility (subject to the design matrix being “almost orthogonal”), our first set of research questions were:

1. What fraction of respondent trials results in a direction reversal, wherein the final Phase III partworths predict that the left (right) concept would be preferred but the respondent actually prefers the right (left)?
2. Given that the ACA model correctly predicts direction (i.e., appropriately calls left or right according to the utility difference predicted by the updated ACA model), do the actual ratings tend to be systematically more extreme than those predicted—do respondents attempt to “fill up” the whole Phase III response scale?

**This decision was based on the results of a previous empirical study involving the same attributes and levels. In that study 97% of the subject-trials respected monotonicity in their subjective ratings of attribute-level acceptability.**
Table 1 summarizes the results of this inquiry. First, we note that more than 25% of the time the updated ACA model does not predict the correct direction; the model predicts left (right) is preferred to right (left) when the actual response indicates the converse. When the ACA model predicts direction correctly, it underpredicts the degree of stated preference ratings almost three-quarters of the time. This result is highly significant ($p < .001$).

**How Well Does the Self-Explicated Model Predict Subjects’ Final Partworths?**

A related question is how well the Phases I/II self-explicated model predicts subject’s final Phase III partworths. Each time a respondent makes a Phase III judgment, ACA uses an OLS regression model to update the partworths. Hence, the partworths are modified in each step, starting from step 1 (in which predictions are based solely on the self-explicated partworths) up through step 15 (in which the effect of the self-explicated partworths may be reduced considerably by information obtained from the Phase III responses).

We computed the correlation between self-explicated partworths and the final Phase III partworths (i.e., those based on Phases I/II and Phase III). This computation was done individual by individual and then summarized in Table 2. We examine the total partworths (across all six attributes) first. As seen from Table 2, the correlations between self-explicated and final partworths average .883. The more important attributes of rent, condition, and safety are associated with even better fits of self-explicated to final partworths.

**How Well Does the Self-Explicated Model Fit the Set of 15 Predictions Step by Step?**

Rather than focusing on only the final partworths associated with the last (fifteenth) paired comparison, we also examined the comparative ability of the self-explicated data alone to predict each of the 15 paired-comparison responses (rather than the final derived partworths). We then compared these findings with those of the updated ACA model that includes the information provided by the paired-comparison responses of all pairs preceding each pair in question. In all, there are 15 predictions; only the first is based strictly on the self-explicated utilities.

Comparative performance of the updated versus self-explicated utilities is judged by two measures—the product moment correlation between actual and predicted and the mean absolute deviation between actual and predicted. Table 3 shows the results. We find relatively high correlations between self-explicated and final partworths, but predicting the actual Phase III response scale values is a different matter. The total sample correlation, averaged over subject trials (15 trials per respondent), is .335. The associated mean absolute error of prediction is 1.808—almost two scale points on the −4 to +4 (transformed) response scale.

The updating model that incorporates all preceding Phase III responses up to that step actually produces a slightly higher mean absolute error of 1.865 (though it is not significant at the $p = .05$ level). However, the correlation of the updated Phase III predictions with the 15 actual responses is higher (.524) than that found for the self-explicated model only (.335). This comparison is statistically significant ($p < .05$). Apparently, signal is provided in the graded paired comparisons over and above the self-explicated predictions, after one allows for a linear transformation of the updated predictions.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>SUMMARY OF RESPONDENTS’ EVALUATIONS OF THE ACA SIMILAR-IN-UTILITY PAIRS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Total sample</strong> (N = 170)</td>
</tr>
<tr>
<td>Fraction prediction wrong direction*</td>
<td>.268</td>
</tr>
<tr>
<td>Standard error</td>
<td>.009</td>
</tr>
<tr>
<td>Fraction in which actual response is more extreme than predicted*</td>
<td>.744</td>
</tr>
<tr>
<td>Standard error</td>
<td>.012</td>
</tr>
</tbody>
</table>

*Given a nonzero actual response.

*Given a correctly predicted direction.

**Table 2**

**SUMMARY OF PRODUCT MOMENT CORRELATIONS BETWEEN SELF-EXPLICATED PARTWORTHS AND FINAL PHASE III PARTWORTHS**

<table>
<thead>
<tr>
<th></th>
<th>Total sample (N = 170)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean overall partworth correlation</strong></td>
<td>.883</td>
</tr>
<tr>
<td><strong>Standard error</strong></td>
<td>.005</td>
</tr>
<tr>
<td><strong>Individual partworth correlations</strong></td>
<td></td>
</tr>
<tr>
<td>Attribute 1 mean (time)</td>
<td>.766</td>
</tr>
<tr>
<td>Standard error</td>
<td>.034</td>
</tr>
<tr>
<td>Attribute 2 mean (noise)</td>
<td>.824</td>
</tr>
<tr>
<td>Standard error</td>
<td>.022</td>
</tr>
<tr>
<td>Attribute 3 mean (safety)</td>
<td>.887</td>
</tr>
<tr>
<td>Standard error</td>
<td>.016</td>
</tr>
<tr>
<td>Attribute 4 mean (condition)</td>
<td>.958</td>
</tr>
<tr>
<td>Standard error</td>
<td>.006</td>
</tr>
<tr>
<td>Attribute 5 mean (size)</td>
<td>.744</td>
</tr>
<tr>
<td>Standard error</td>
<td>.034</td>
</tr>
<tr>
<td>Attribute 6 mean (rent)</td>
<td>.949</td>
</tr>
<tr>
<td>Standard error</td>
<td>.011</td>
</tr>
</tbody>
</table>

*In general, the information provided by the graded paired comparisons is insufficient for parameter estimation without inclusion of the self-explicated response data. Hence, one cannot estimate (in general) partworth parameters from Phase III paired-comparison responses alone.
Table 3
PERFORMANCE OF SELF-EXPPLICATED VERSUS (UPDATED) MODEL ACROSS ALL 15 STEP-BY-STEP PREDICTIONS: CORRELATIONS AND MEAN ABSOLUTE ERRORS

<table>
<thead>
<tr>
<th>Total sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N = 170 )</td>
</tr>
</tbody>
</table>

| Correlations based on self-explicated alone | .335 |
| Standard error | .012 |
| Correlations based on updated model | .524 |
| Standard error | .015 |
| Mean absolute error of self-explicated alone | 1.808 |
| Standard error | .037 |
| Mean absolute error of updated model | 1.865 |
| Standard error | .041 |

Holdout Sample Comparison

How strong is this signal? Working with the same data used here and the two sets of holdout full profiles, Agarwal and Green (1989) first compared three sets of predictions: (1) the ACA self-explicated model alone, (2) the updated ACA model based on only the first eight Phase III paired comparisons, and (3) the updated model based on all 15 paired comparisons. The first three columns of Table 4 give the results.

In the case of correlations between model and holdout preferences, the gain in correlation is about five to eight points for the fully updated ACA model versus the self-explicated model alone. The gain in going from the first eight pairs to all 15 pairs is one to two correlation points. Corresponding improvements in predicting first-choice hits and correct rank position are much less.3

Bearing in mind that the fully updated model entails 15 additional datapoints, we note that there appears to be considerable noise in subjects’ paired comparisons. The relatively high mean absolute errors of Table 3 support this conclusion whereas Table 1 shows the bias that results as subjects attempt to utilize the full range of the Phase III scale.

Possibly Phase III contributions to prediction would be even less if the heuristic were used on paired comparisons between partial profiles composed of three to five attributes (the maximum included in the ACA model). Not only would each comparison be more complicated to evaluate, but the heuristic should be able to come up with even closer (and, possibly, more difficult to judge) pairs.

The Pencil-and-Paper Exercise

In Agarwal and Green’s empirical analysis of the dataset used here, each of the 170 subjects was given a pencil-and-paper questionnaire and asked to perform two tasks:

1. Rate the acceptability of each attribute level (three levels per attribute) on an 11-point 0 to 10 equal-interval scale, anchored at 0 = completely unacceptable to 10 = completely acceptable.

2. Allocate 100 points across the six attributes so as to reflect their relative importance in evaluating apartment profiles.

Table 4
CROSS-VALIDATION PERFORMANCE OF FOUR PREDICTION MODELS ON TWO SETS OF HOLDOUT PROFILES
(Source: Agarwal and Green 1989)

<table>
<thead>
<tr>
<th>Correlations</th>
<th>ACA self-explicated alone</th>
<th>Updated ACA</th>
<th>Updated ACA</th>
<th>Pencil/paper exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(first 8 pairs)</td>
<td>(all 15 pairs)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 profiles</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Standard error</td>
</tr>
<tr>
<td>16 profiles</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Standard error</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>First choice hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 profiles</td>
</tr>
<tr>
<td>Mean (%)</td>
</tr>
<tr>
<td>Standard error</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank position hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 profiles</td>
</tr>
<tr>
<td>Mean (%)</td>
</tr>
<tr>
<td>Standard error</td>
</tr>
<tr>
<td>16 profiles</td>
</tr>
<tr>
<td>Mean (%)</td>
</tr>
<tr>
<td>Standard error</td>
</tr>
</tbody>
</table>

3 No results are given for the 18 holdout profile case because first-choice hits were not able to discriminate among models.
Each raw acceptability scale was left as is; the resulting scale was not translated or stretched to obtain endpoints of 0 and 1.0. The constant-sum points were normalized (for convenience) by dividing each original entry by 100.

A simple compositional model was constructed by finding the sum of the products of acceptability scores by importance for all relevant attribute levels. Each holdout profile in the Agarwal and Green study entailed a 0–100 rating on a likelihood-of-renting scale; these (orthogonally designed) profiles were the same holdout profiles used in assessing ACA’s predictive validity (see the last column of Table 4 for pencil-and-paper results).

The pencil-and-paper compositional model out-predicted the full ACA model on all three criterion measures—correlation, first-choice hits, and rank position hits. As noted from Table 4, though the pencil-and-paper self-explicated results are not dramatically higher than the results of the fully updated ACA models, their improvement is impressive if one realizes that the respondent is not subjected to 15 additional paired comparisons—quite a saving in interview time.

CAUTIONS AND CAVEATS IN USING ACA

In our view, the ACA partworth estimation process has several weaknesses.

1. Phase I attribute-level scales mandate equal subjective intervals and equal ranges across attributes by forcing each scale to range from 0 to 1.0, with all intermediate preferences equally spaced.

2. Phase II attribute importance ratings are too coarse; only four response values (4 = extremely important; 1 = not important at all) are permitted.

3. The Phase III paired-comparison evaluation of partial profiles is both difficult to do and unrealistic. It is difficult to do because the ACA heuristic attempts to make all pairs essentially indifferent yet provides a relatively broad (9-point) scale for respondents to use. It is unrealistic because real-world options rarely vary on only two or three attributes at a time. (However, larger numbers of attributes can entail greater information overload.)

4. Phase III comparisons present a scale compatibility challenge. For proper estimation, scales should be commensurate between Phases I/II and Phase III. Because question wordings and measurement scales differ so much between phases, it seems virtually impossible to attain comparability of response units.

Some Constructive Suggestions for Improving ACA

In the spirit of improving an already popular software package, we suggest that Sawtooth consider the following modifications of ACA.

1. Include, as an option, the use of finer grained acceptability scales (e.g., a 0–10 scale of acceptability) that are not translated and stretched to endpoints of 0 and 1.0 and, in particular, are not transformed to equal intervals along the scale.

2. Include a finer grained importance scale (e.g., 1–10) that would provide the subject a greater range of response. By the use of cursor arrow keys, one also could consider

a more or less continuous scale ranging (say) from 1 to 10.

These two simple options for Phase I and Phase II, respectively, should increase the predictive accuracy of ACA’s self-explicated section. (See the last column results in Table 4 as an illustration.)

For the Phase III data source, we assume that Sawtooth does not want to eliminate or replace its traditional mainstay—the graded paired-comparisons task. Assuming that this task stays, we suggest the following possible options.

1. An option for selecting paired comparisons on some basis other than that assumed by the equal-in-utility heuristic. Subject to maintaining attribute and level balance, one could select pairs randomly. Alternatively, one could select pairs that make the design matrix as orthogonal as possible.5

2. An option for including a partworth updating feature that does not require commensurate units between Phases I/II and Phase III. For example, some type of hybrid model updating (see Green 1984) could be employed in which stagewise regression is applied to residuals obtained from a linear or nonlinear function of the self-explicated partworths.

Version 3.0 of ACA allows the user carte blanche in choosing a Phase III scale that can range from 1 to 2, 1 to 3, . . . to 1 to 9. In our view this innovation, if anything, exacerbates the problem of trying to find commensurate units between Phases I/II and Phase III. Clearly, not all of these values can be correct in a given problem.

A Formal Procedure for Finding Commensurate Scale Units Between Phases I/II and Phase III

Our final suggestion for improving ACA directly addresses the problem of compatibility between the Phases I/II model of equation 2 and the Phase III model of equation 3. The suggested procedure also is appropriate for version 3.0 of ACA, in which the user chooses the Phase III scale range.

We have suggested that the scores given to the paired comparisons are more dispersed than one would predict from the partworths computed from the self-explicated model and the paired comparisons up to that point.9 A behavioral reason (suggested previously) is that the pairs chosen by ACA are such that the predicted score is de-

5Empirical research would be needed, however, to show whether these suggestions lead to higher cross-validation of holdout profiles than ACA’s current pair-selection procedure.

9One might argue that the reason for the variance across predictions being less than the variance across actual values is that the variance of a predicted value is typically smaller than the variance of an actual value (as in regression). In our setting, however, we found algebraically that the variance of the predicted values (based on the assumed OLS model) is greater than the variance of the actual values. (A formal note on this matter is available from the authors.)
signed to be near zero. In evaluating a paired comparison, however, a respondent has a tendency to use more of the (transformed) \(-4\) to \(+4\) scale.

This observation suggests that if the actual scores are linearly transformed before estimation of the partworths, there could be an improvement over the original paired-comparisons results. Formally, we can characterize the problem by the following regression equations:

\begin{align}
\mathbf{y} &= \mathbf{I} \mathbf{\beta} + \mathbf{e} \\
\mathbf{v} &= \mathbf{X} \mathbf{\beta} + \mathbf{\delta}
\end{align}

where \(\mathbf{\beta}\) denotes the partworths, \(\mathbf{y}\) the self-explicated scores, and \(\mathbf{v}\) the transformed scores from the paired comparisons. We are given:

1. \(\mathbf{I}\), the identity matrix of dimensions equal to the total number of levels across all attributes.
2. \(\mathbf{X}\), a matrix of entries \(-1, 0, 1\); row \(i\) has entries \(-1(\pm1)\) whenever a level in the partial product of the \(r\)th paired comparison appears on the left (right) side of the screen and 0 if a level does not appear at all.

We do not observe \(\mathbf{v}\) but rather \(\mathbf{z}\) where

\[\mathbf{v} = \mathbf{\mu} + \mathbf{\alpha} \mathbf{z}\]

The problem, then, is to estimate \(\mathbf{\beta}\), where we know \(\mathbf{X}\), \(\mathbf{y}\), and \(\mathbf{z}\); \(\mathbf{\mu}\) and \(\mathbf{\alpha}\) are nuisance parameters in this formulation.

The partworths \(\mathbf{\beta}\) and, coincidentally, the transformation parameters \(\mathbf{\mu}\) and \(\mathbf{\alpha}\) can be estimated by alternating least squares (ALS). The steps are:

1. Given \(\hat{\mathbf{\beta}}_{00}\), \(\hat{\mathbf{\mu}}_0\) and \(\hat{\mathbf{\alpha}}_{00}\) can be estimated from simple OLS regression where the dependent variable is \(\mathbf{X} \hat{\mathbf{\beta}}_{00}\) and the independent variable is \(\mathbf{z}\).
2. Given \(\hat{\mathbf{\mu}}_0\) and \(\hat{\mathbf{\alpha}}_{00}\), \(\hat{\mathbf{\beta}}_{r+1}\) can be estimated from equations 4 and 5 by replacing \(\mathbf{v}\) in 5 by \(\hat{\mathbf{v}}_{00} + \hat{\mathbf{\beta}}_{00} \mathbf{z}\). An initial value \(\hat{\mathbf{\beta}}_{00}\) for the partworths can be obtained from the self-explicated model to start the whole iterative process.

This suggested transformation keeps ACA's current updating process essentially intact while appropriately addressing the problem of scale units' commensurability.

CONCLUSIONS

As Johnson (1987b) has demonstrated via Monté Carlo analysis, the Phase III paired-comparisons section actually could add enough noise that ACA's Phases I/II data predict holdout profiles better than the full model that includes Phase III responses to the pairs. Fortunately, in the application reported here, such was not the case. However, we found that the finer grained Phase I/II scales employed in the pencil-and-paper exercise outperformed ACA's full model on two sets of holdout profiles. (As described previously, changes in Phase I/II's scales appear to be a straightforward computer programming task.)

Scale compatibility between Phases I/II and Phase III is a more complex matter. We offer suggestions for changing the current partial profile selection and parameter estimation procedure, but further empirical testing of the alternatives is necessary.

ACA is a highly visible and popular procedure for collecting conjoint data. It is versatile, easy to learn, and easy to use. However, we believe that its more formal features can be improved substantially without detracting from its cosmetically attractive qualities.

We hope the suggestions given here are constructive. Sawtooth Software has maintained an open mind to user suggestions for improving its ACA package. Our comments are offered in response to the firm's expressed desire to solicit software changes that show potential for improving ACA's reliability and versatility.

REFERENCES


