Sawtooth Software

RESEARCH PAPER SERIES

Accuracy of Utility Estimation in ACA

Richard M. Johnson

April 6, 1987

530 W. Fir St.
Sequim, WA 98382
(360) 681-2300
www.sawtoothsoftware.com

Accuracy of Utility Estimation in ACA

Richard M. Johnson
Sawtooth Software, Inc.
April 6, 1987

ACA users have asked about the accuracy of utility estimation under various questionnaire conditions. This report describes a Monte Carlo study that answers some of those questions.

Questions about accuracy of estimation can be answered by comparing estimated utilities with the underlying true values. An experimental version of the ACA questionnaire program was produced to answer questions automatically, based on known utilities. The estimated utilities were compared to the true values by computing correlation coefficients.

It should be noted that this approach doesn't tell us anything about "human factors" issues such as the conditions likely to minimize respondent fatigue and confusion. Those problems may be even more important than the statistical issues treated here.

Study Plan:

An experimental version of the ACA questionnaire program was developed to do these things:

1. Construct a random true set of utilities:

    a. rectangularly distributed in the range +2.0 to -2.0. (Some choice of scale had to be made. Since ACA uses importance questions with answers in the range of 1 to 4 it seemed natural to construct utilities having approximately the same range.)

    b. with means of zero for the levels of each attribute.

2. Answer questions automatically:

    a. for rank order questions, use the rank orders of the true utilities.

    b. for importances, use the range of true utilities for that attribute, rounded to the nearest integer.

    c. for paired questions, use the difference in sums of true utilities for the two concepts, after rounding to the nearest integer and adding 5.

3. Compute the squared correlation coefficient between true and estimated utilities for each interview.

We compared results obtained by different questionnaire treatments, including:

1. Number of pairs
2. Pairs consisting of 2 vs 3 vs 4 vs 5 elements
3. Different levels of respondent error
4. Prohibiting reversals of a priori order relations

Each statistic reported is an average for 30 simulated interviews. All interviews used the same numbers of attributes and levels. There were a total of eight attributes and 31 levels. The numbers of levels per attribute were 5, 5, 5, 5, 3, 3, 3, and 2.

Results:

The square of a correlation coefficient can be interpreted as the percentage of variance in one measure which is accounted for by the other. Its complement is the proportion of variance which is not accounted for, or "relative error variance." We are concerned primarily with error variances, so we report percent error variances as well as squared correlations.

Accuracy and the Number of Pairs. The first results show relative errors of estimation for various numbers of pairs, each containing just two attributes.

Table 1.
Accuracy and Number of Pairs

Number of Pairs	Squared Corre-lation	Percent Error Variance
0	.884	11.6
10	.931	6.9
20	.958	4.2
30	.976	2.4
40	.985	1.5
50	.988	1.2
100	.994	0.6
500	.999	0.1

The first row shows how good an estimate of the true utilities can be made from just the rank order and importance questions asked at the beginning of the interview.

For up to 50 pairs, each 10 pairs decrease the error level by about a third.

Although no questionnaire would ever have as many as 100 pairs, the last two lines are presented just to show that as the number of pairs increases the estimated utilities converge to the true values, indicating that the ACA estimation process is unbiased.

For this number of attributes and levels, the number of pairs that would automatically be asked by ACA is 35. If the respondent were to answer each question with perfect consistency, as in this simulation, we would expect his estimated utilities to have a squared correlation with his true utilities of about .98, for a relative error variance of about 2%.

Accuracy and Types of Pairs. The next question has to do with pairs composed of 2, 3, 4, or 5 attributes. For each of those conditions 30 interviews were simulated, each interview containing 30 pair questions. The results were as follows:

Table 2.
Accuracy and Types of Pairs

Number of Attributes Per Pair	Squared Corre-ration	Percent Error Variance
2	.976	2.4
3	.984	1.6
4	.986	1.4
5	.986	1.4

The case of 2 attributes per pair is repeated from Table 1. The other cases provide new information. There is relatively little difference among these treatments. Since the 2 and 3 attribute cases are much easier for respondents, there seems to be little benefit from showing pairs based on more than 2 or at most 3 attributes.

Response Error. All the results above assume responses of perfect consistency, although we can be sure this will not happen with human respondents. Table 3 shows the effects of random response error.

In ACA a respondent expresses his preferences for pairs on a 1 to 9 scale, with a response of 1 indicating strong preference for one alternative, 9 indicating strong preference for the other alternative, and 5 indicating no preference. In the simulations reported above we computed the true utility difference for the alternatives, rounded it to the nearest integer, added 5, and took the result as the response. The only random error was that involved in rounding to integer values. In these computations we also add an additional random component before rounding. This component has an average of zero, and is rectangularly distributed with a maximum absolute value of 0.5, 1.0, or 2.0.

ACA is quite efficient at choosing questions with alternatives having nearly equal utilities, and the true difference in utilities for the two alternative concepts averages less than a scale point. Adding a random component as large as plus or minus 2.0 is likely to "swamp" the signal with noise, and we should expect the estimation to be quite poor. Table 3 reports results for three levels of random response error, and also repeats result. for the error-free case.

Table 3.
Accuracy and Level of Random Response Error
(Percent Error Variances)

Number Of Pairs	Maximum Additional Random Error			
	none	.5	1.0	2.0
0	11.6	11.6	11.6	11.6
10	6.9	7.5	8.5	13.2
20	4.2	5.2	7.1	15.2
30	2.4	3.1	5.9	15.4
40	1.5	2.2	4.3	12.5
50	1.2	2.0	3.9	11.2

The first column of the table is repeated from Table 1. The other three columns present new information. For each column we also repeat the relative error variance resulting from an initial estimate based on error-free preference rankings and importance ratings.

For random response errors in the range of plus or minus a half scale point, the error of estimation increases only moderately. Random response errors in the range of plus or minus a full scale point lead to larger but still tolerable errors of estimation.

However, random response errors in the range of plus or minus two scale points produce dramatically larger errors of estimation. In fact, with this level of response error, data from the pairs actually degrade initial estimates made from (error-free) rank orders and importance ratings unless there are as many as 50 pairs. This curious result is not of much importance because we would seldom get error-free initial responses from a respondent who was this inconsistent in the pairs.

Order Restrictions. Our last point is concerned with insuring that estimated utilities maintain the same rank order, within each attribute, as the initial rank orders. For some attributes we can be sure that the respondent should prefer one level to another, such as preferring "more durability" to "less durability," or "lower price" to "higher price." With high levels of response error, estimated utilities may violate such order restrictions. In that case we could either leave them alone, hoping they will be straightened out later, or make corrections "on the fly." The correction used here was to replace utilities that violate initial rank orders with their averages, thus tying them.

Table 5.
Accuracy And Order Restrictions for the
Highest Level of Random Response Error
(Percent Error Variances)

Number Of Pairs	No Corrections	Corrections Made
10	13.2	12.6
20	15.2	13.8
30	15.4	13.7
40	12.5	11.0
50	11.2	9.5

Table 5 shows that in the case of very noisy data, there is a modest increase in accuracy if estimated utilities are not permitted to have order reversals within attributes. Restrictions were also examined for lower levels of response error, and they were not found to have appreciable effect in those cases.

ACA does not currently enforce such restrictions, but perhaps it should. However, the corrections are not without cost, since they destroy the least squares properties of the estimated utilities.

Summary:

ACA appears to be quite accurate at estimating respondents' utilities, provided they respond consistently. In this moderate-sized example (eight attributes with 31 levels), the average squared correlation between actual utilities and estimates based on just the "front end" questions, with no pairs, is about .88. With this number of attributes and levels ACA will normally ask 35 pair questions. With 35 pairs the average squared correlation between actual and estimated utilities is about .98, assuming no response error.

There is some increase in accuracy for pairs composed of more than two attributes, but the gain is so slight that the risk of respondent confusion probably more than offsets it. Under most circumstances pairs should not be allowed to differ in more than three attributes.

With moderate levels of response error the accuracy of estimation still appears acceptable. However, if the random component becomes much larger than the actual differences in utilities between the two alternatives being offered, estimation is badly degraded.

Finally, there is some support, with very noisy data, for the practice of checking to insure that estimated utilities do not violate a priori order relations, and undoing those violations by forcing ties where necessary. ACA does not do this at present but the addition of this feature is under consideration.