



Sawtooth Software

TECHNICAL PAPER SERIES

Conjoint Value Analysis (CVA) Version 3.0

CVA

A Full-Profile Conjoint Analysis System From Sawtooth Software

Version 3



Copyright Sawtooth Software, Inc.
Sequim, Washington USA (360) 681-2300
<http://www.sawtoothsoftware.com>

General Description

The CVA System is a complete software package for traditional full-profile conjoint analysis. CVA is a component within Sawtooth Software's SMRT system. SMRT stands for Sawtooth Software Market Research Tools. CVA includes:

- A *designer* for creating the stimuli (product profiles) displayed in conjoint interviews
- A *questionnaire component* for producing computer-based interviews or paper-and-pencil surveys
- Two *utility calculators* for estimating each respondent's part worth utilities (strengths of preference) for product features
- A *market simulator* for testing alternative product scenarios and displaying average utilities and importances.

CVA stands for "Conjoint Value Analysis." The CVA System is one of three conjoint software packages from Sawtooth Software, all available as components within SMRT. The other two are ACA (Adaptive Conjoint Analysis) and CBC (Choice-Based Conjoint). We have designed each package to bring unique advantages to different research situations.

About Conjoint Analysis

Conjoint analysis is useful for learning how potential buyers of a product or service value its various aspects or features. Conjoint analysis is equally useful for studying products and services (for convenience, we use the word "product" in both cases).

Examples of questions that lead market researchers to use conjoint analysis are:

- Which would be more valuable to the customer: longer warranty or more product performance?
- What would be the effect of raising our price 10%?
- What would happen if we added another product to our line that had 20% less performance but a 30% lower price?

When using conjoint analysis, a product is thought of as being made up of various **attributes** (sometimes called "factors"). For example, automobile attributes may be: manufacturer, color, type of engine, number of doors, and price.

Each attribute has several possible **levels**. For example, the levels of "number of doors" may be: two, four, and five (to include station wagons). The levels of "engine type" may be: 4 cylinder, 6 cylinder, and 8 cylinder. (The word "feature" is often used in place of "attribute level.")

Conjoint data are collected from survey respondents by asking about their preferences for hypothetical product concepts, described in terms of specific attributes and levels. In analyzing the data, one infers the value to each respondent of having each possible attribute level. These estimated values are called **part worths** or **utilities**. CVA's calculation routines can estimate each respondent's part worth utility values for each level of each attribute from data collected by the conjoint questionnaire. These values are then used to make predictions about how respondents would choose among new or modified products. The goal is to conclude what product changes would have the most beneficial (or least detrimental) effect on share of preference, or which would maximize the likelihood that buyers would choose specific products.

In conjoint analysis, product concepts can be shown to respondents one at a time, or they can be presented in pairs. *Pairwise* presentation can be harder for the respondent, because each question requires understanding two concepts rather than just one. However, the comparative nature of the pairwise task may let the respondent make finer distinctions and contribute more information than *single concept* (popularly called *card sort*) presentation.

The term *full profile* refers to product concepts that are described with respect to *all* the attributes being studied. For example, if the products in a particular category had ten important attributes, a full profile product concept would describe the product on all ten attributes.

CVA produces either single concept or pairwise full profile designs. It is questionable that respondents can deal with concepts having more than about six attributes. However, over the years we occasionally have heard of people wanting to use CVA's experimental design program for applications that we had not envisioned. To give users this flexibility, and because modern PCs offer processing speeds and memory not available before, we have decided to allow this newest version of CVA to support up to 30 attributes, each with up to 15 levels. Despite this capacity, we recommend that the number of attributes and levels for studies in practice be much smaller than these limits. With more than about six attributes, the questions may be too complex for most respondents to handle, and the number of questions that must be asked of each respondent may become excessive. If more attributes and levels are necessary, it can be advantageous to show respondents pairs of concepts described on subsets of the attributes, as in our ACA System for Adaptive Conjoint Analysis.

About CVA

CVA features a variety of tools and functionality. Not all of these tools need to be used in a particular project.

The authoring and design tools can be used to produce computer-administered or paper-and-pencil questionnaires. We do not recommend the questionnaire design tool be used for studies with many attributes. With too many attributes, the questions are likely to provide too much information to the respondent, who may become overloaded and confused. For studies with more than about six attributes not dealing specifically with pricing research, we often recommend using the ACA System for Adaptive Conjoint Analysis.

The calculation routines can be used to analyze questionnaire data and calculate respondent utilities. The two standard methods are "Ordinary Least Squares" (OLS) regression and "monotone" (nonmetric) regression. The OLS routine is appropriate for ratings-based data, and the monotone program for rankings-based data. An optional advanced module called CVA/HB is available for hierarchical Bayes (HB) estimation.

SMRT's market simulator is shared across our three conjoint systems. The market simulator lets the researcher specify a market consisting of several products defined in terms of attribute levels, and then estimate the shares of preference that each product would receive. If you have used single-concept presentation (described later) and purchase likelihood ratings, then the market simulator can be used to estimate purchase likelihood for product profiles.

Question Formats

Pairwise Presentation

The default questionnaire layout in CVA is Pairwise Comparison. With pairwise presentation, respondents are asked to compare two products, as shown below:

Which of these two luxury skyboxes would you prefer?

12-person seating capacity On about the 30-yard line \$15,000 per year	or	18-person seating capacity On about the 20-yard line \$22,500 per year						
Strongly Prefer Left	Somewhat Prefer Left	Indifferent	Somewhat Prefer Right	Strongly Prefer Right				
1	2	3	4	5	6	7	8	9

With small numbers of attributes, the paired comparison format has desirable characteristics. Paired comparisons have a long history in the field of psychometrics, where they have been the method of choice for the precise measurement of small subjective differences. When respondents compare two products side-by-side, it can help them draw finer distinctions between products offering different features. However, if two products are shown per question, respondents must evaluate twice as much information as single concept designs (described below). All research we are aware of to date suggests that relative part worths are very similar whether estimated from single concept or paired comparison conjoint questionnaire formats.

We suggest the pairwise design for most conjoint projects. However, the pairwise questionnaire only captures the relative differences in a respondent's preferences for attribute levels. Pairwise questions never measure respondents' absolute level of interest in the product concepts. Comparative information is all that is necessary for estimating relative part worths and running competitive market simulations. If you need to run "purchase likelihood" simulations within the market simulator, you should use the single concept presentation, described next.

Single Concept Presentation

Single concept presentation is commonly known as “card-sort.” Respondents can be shown one product at a time and asked to rate how likely they would be to purchase each. An example is shown below:

How likely are you to buy a 2-year lease for the following skybox at the Wendell State stadium?

Slide the pointer to the position on the scale to indicate your answer. A "0" means you definitely would NOT buy this lease, and a "100" means you definitely would lease this skybox.

**18-person seating
capacity**

**On about the
20-yard line**

\$22,500 per year

Definitely
Would Not
Buy

Definitely
Would Buy

50

This method is particularly useful if the “purchase likelihood model” is to be used in the simulator. CVA’s ordinary least squares (OLS) routine is appropriate for analyzing single concept ratings.

Alternatively, each of a set of product profiles can be printed on individual cards, and respondents can be asked to sort the cards from most to least preferred. In this case, the responses are rank orders of preference, and CVA’s monotone regression routine is suggested for developing utilities.

Evidence to date suggests that either questionnaire technique (pairwise or single-concept) produces quite similar part worths and importances.

Advantages and Challenges of CVA Questionnaires

Although CVA's question format has advantages in studies with small numbers of attributes, its task can be challenging for the respondent. To be sure that your CVA questionnaire is appropriate for the study, you must:

- keep the number of attributes small,
- pretest the questionnaire, and
- conclude that the resulting utilities are reasonable,

before you proceed.

Any conjoint questionnaire can be difficult for the respondent, but CVA questionnaires can be particularly so. Unlike our ACA System for Adaptive Conjoint Analysis, CVA does not use any strategy for simplifying the respondent's task. Every question involves consideration of all the conjoint attributes. CVA can create either single concept (card sort) or pairwise comparison (two concepts at a time). With pairwise questions, the amount of information the respondent must consider is twice as great as with single concept methods.

Unless the number of attributes is quite small, (about six or fewer) there is risk that respondents will be overloaded, and that the quality of the data collected may be poor. For this reason, we recommend other types of questionnaires if many attributes must be studied.

In summary, although the CVA question format can be very effective when the number of attributes is not too large, we are concerned that inappropriate use of CVA may produce questionnaires that are too difficult for respondents. As with any conjoint method, but particularly so in this case, it is essential to pretest the questionnaire:

- answer the questionnaire yourself, and analyze your own data to make sure the utilities mirror your own values.
- have others answer the questionnaire to report to you about whether it's too difficult.
- have a sample of relevant respondents answer the questionnaire, and analyze their data. Be particularly alert for "nonsense" results with respect to price. If many respondents have higher utilities for higher prices, then the data are suspect.

If you do keep the number of attributes small, and if you do pretest the questionnaire and conclude that the resulting utilities are reasonable, then you can proceed with confidence.

Selecting the Number of Tasks

After you have defined attributes and levels and decided which questionnaire format is best for your study, you need to decide how many conjoint questions (tasks) to ask. This decision depends on how many separate part worth parameters need to be estimated. The more attributes and levels in your study, the more part worths to be estimated; and the more parameters to estimate, the longer the questionnaire.

CVA examines the list of attributes and levels in your study and provides an initial recommendation regarding the number of tasks (conjoint questions) to ask. The recommended number of tasks provides three times the number of observations as the number of parameters to be estimated. The number of parameters to be estimated is determined by the formula:

$$\text{Total number of levels} - \text{number of attributes} + 1$$

Although it is mathematically possible to estimate utilities by asking only as many questions as the number of parameters to be estimated, we strongly recommend against this practice. Asking the recommended number of tasks helps ensure enough information to calculate stable estimates for each respondent.

If your design includes many attributes and levels, you may notice that the recommended number of questions exceeds what you could reasonably expect respondents to complete. This leads to a difficult decision. Too few questions limit the degrees of freedom of the design; but if you ask too many conjoint questions respondents may become fatigued or bored and provide data of poor quality. In our experience, asking more than about 30 conjoint questions may result in poor quality data. Testing your questionnaire with actual respondents can help you determine the optimal number of questions to ask for your project.

CVA will not let you specify fewer questions than the number of parameters to be estimated, and will warn you if you don't ask at least 1.5x the number of parameters to be estimated. A number of studies in the literature suggest that some experienced conjoint analysts are willing to ask as few as 1.5x the number of questions as parameters to estimate. Indeed, the use of HB estimation may slightly reduce the number of questions you decide to ask respondents, as it estimates part worths based on information from the current respondent plus information from the other respondents in the same dataset.

Generating CVA Designs

After you have specified attributes, any assumed *a priori* ordering for levels within attributes, chosen the questionnaire format, how many tasks to ask, and selected whether to discard "obvious" tasks, you are ready to generate a design. The design specifies the combination of attribute levels (profiles) shown in each conjoint question.

Orthogonality and CVA Designs

The conjoint design is a critical component to the success of any conjoint project. Attributes must vary independently of each other to allow efficient estimation of utilities. A design with zero correlation between pairs of attributes is termed "orthogonal." Level balance occurs if each level within an attribute is shown an equal number of times. Designs that are orthogonal and balanced are optimally efficient.

In the real world, it might not be possible to create a perfectly balanced, orthogonal design for a particular set of attributes and prohibitions consisting of a reasonable number of tasks. The CVA approach produces high quality designs automatically, although they probably will not be perfectly balanced or orthogonal. CVA lets you test the efficiency of a design before fielding your study. Testing the design also lets you study the impact of including prohibitions, or asking fewer than the recommended number of questions.

CVA provides an easy-to-use tool for generating well-balanced, "nearly-orthogonal" designs. CVA generates a pool of potential conjoint questions (from which the final design will be chosen) using a relatively simple procedure. For each question to be composed, it does the following: For each attribute it picks a pair of levels randomly from among all permitted pairs that have been presented the fewest times. A random decision is made about which levels will appear on the left and on the right (for pairwise designs). No pair of levels will be repeated until all other permitted pairs have been shown. Each pair of levels will be shown approximately the same number of times, and each level from one attribute is equally likely to be shown with any level from another attribute.

CVA's designer uses the following steps to select efficient designs, given the questionnaire specifications:

- 1) CVA generates a pool of potential conjoint questions equal to, by default, a pool multiplier of 10 times the requested number of questions (assuming that many unique questions exist).
- 2) The D-efficiency of the design is calculated for the pool, excluding one conjoint question at a time (only using the first 10x tasks created, where x is the requested number of

tasks). The one task that contributes least to the efficiency of the design is discarded, and the process repeated until the desired number of tasks remains.

- 3) CVA then examines every potential 2-way swap of conjoint questions that remain with those that were discarded or are available in the pool of potential conjoint questions. CVA swaps any pairs of questions that result in an increased efficiency.
- 4) Next, CVA examines the frequency of level occurrences for each attribute. It investigates interchanging levels that are under-represented in the design with levels of the same attribute that are over-represented. Any changes that result in improved D-Efficiency are retained.
- 5) Finally, for pairwise designs, CVA flips left and right concepts to improve the left/right balance of the design.

CVA repeats steps 1 through 5 ten separate times and chooses the best final solution of the ten (you can over-ride CVA's default of ten passes to enable CVA to potentially find a slightly better design). For studies with many attributes and levels, a pool multiplier of 100 or more usually improves CVA's ability to find efficient designs more quickly.

CVA uses the final solution within the questionnaire. Every respondent receives the same set of questions in CVA surveys. You can export the design specifications to a comma-delimited text file if you wish. You can also import designs for use within CVA from comma-delimited text files. This enables advanced users to further customize or modify the design (for example, to add a few user-specified holdout cards).

D-efficiency

CVA's iterative designer seeks to maximize D-efficiency, given the conditions and constraints defined by the user. D-efficiency is described in the article by Kuhfeld, Tobias, and Garratt (1994), "Efficient Experimental Design with Marketing Research Applications," *Journal of Marketing Research*, 31 (November), 545-557.

Paraphrasing the article, D-efficiency is a function of the determinant of the $X'X$ matrix, given by the formula:

$$\frac{100}{N_D |(X'X)^{-1}|^{1/p}}$$

where

N_D	=	number of tasks
p	=	number of attributes
X	=	the design matrix using orthogonal coding

If a design is orthogonal and balanced (each level within an attribute shown an equal number of times), then it has optimum efficiency. The D-efficiency measures the goodness of the design relative to the hypothetical orthogonal design. A perfect design will be both orthogonal and balanced and will result in an efficiency of 100. However, an orthogonal design is not always possible, given the number of attribute levels and requested number of tasks. A final efficiency of less than 100 may still indicate a satisfactory design.

Attribute Prohibitions

With both questionnaire formats (single-concept or pairwise), the researcher can specify that particular pairs of levels should not be shown at the same time. CVA permits both across- and within-attribute

prohibitions. In general, we caution against using prohibited pairs since they usually have a significant negative impact on the efficiency of your design. Prohibitions should be used only to eliminate combinations that respondents would recognize as impossible or absurd. We caution against prohibiting combinations of attributes if the only reason is that they are unlikely or don't exist yet in the market.

Questionnaire Design Example

The example below shows the results of the first of ten passes for a hypothetical conjoint design. After showing CVA's output, we'll discuss it in some detail.

```
Begin Pass #1
# of tasks thrown out because of duplicates or obvious = 29
Candidate set successfully built with 100 tasks.
D-efficiency with...
99 tasks = 90.077068
98 tasks = 90.328815
97 tasks = 90.568669
96 tasks = 90.787874
95 tasks = 90.977933
94 tasks = 91.164087
93 tasks = 91.350241
92 tasks = 91.528866
91 tasks = 91.711268
90 tasks = 91.880740
.
.
(steps 31 to 89 excluded for the sake of brevity)
.
.
30 tasks = 95.680526
29 tasks = 95.574521
28 tasks = 95.562191
27 tasks = 95.579438
26 tasks = 95.603218
25 tasks = 95.605327
24 tasks = 95.595797
23 tasks = 95.679198
22 tasks = 95.519393
21 tasks = 95.435458
20 tasks = 95.125459
19 tasks = 94.922623
18 tasks = 94.515545
17 tasks = 94.214803
16 tasks = 94.058042
15 tasks = 93.771466
14 tasks = 92.878482
13 tasks = 91.170233
12 tasks = 89.469304
11 tasks = 87.844999
10 tasks = 84.818371
# of "thrown out" tasks substituted back into design = 2
# of levels changed to improve balance = 1

D-EFFICIENCY for pass # 1 = 86.948371
Pass 1 saved.
```

In the example above, we specified that we wanted CVA to create a questionnaire with just 10 tasks, or the minimum number of questions to permit estimation of utilities for this design. In practice, this would be less than optimal, as we will explain.

The example design had 4 attributes, with 5, 3, 3, and 2 levels. The first three had *a priori* levels, and two prohibitions were specified. The total number of parameters to be estimated is given by the formula: (total #levels - #attributes) + 1. For this example it is $(13-4) + 1 = 10$.

CVA begins by generating a pool of 100 conjoint questions (ten times the ten questions we requested). The quality of the design is reported at each iteration as D-efficiency, which is bounded by 0 and 100. The first iteration in the example above shows the D-efficiency after discarding the one question (task) which contributes least to the design:

99 tasks = 90.077068

After the last iteration and final adjustments to the design, the D-efficiency for the final design is displayed for the 10-task design: 86.948371.

The efficiency is recalculated each time a conjoint profile is discarded. Notice that the quality of the design increases at first as inefficient questions are discarded, and then begins to decrease as the design thins out. After only about 14 tasks remain, efficiency decreases more rapidly as the last few are discarded. With 10 tasks, the efficiency is at its lowest point, 84.818371.

CVA reports that two questions that had been discarded were substituted back into the design, and that levels have been changed for improved balance once. Twenty-nine tasks were discarded because they were duplicates or “obvious,” so we know CVA actually generated 129 tasks in the process of creating a design pool with 100 unique tasks. Finally, a message appears that the results for Pass 1 have been saved.

We have stressed the importance of including enough conjoint questions to ensure efficient designs. However, design efficiency is not the only reason for including two to three times as many questions as parameters to be estimated. All real-world respondents answer conjoint questions with some degree of error, so those observations beyond the minimum required to permit utility estimation are needed to refine and stabilize utility estimates.

While there is generally a positive relationship between the number of conjoint tasks in the design and D-efficiency, there are many exceptions. It is possible for designs with few or no degrees of freedom to have 100% D-efficiency. This means that this design is optimally efficient for estimating main-effect parameters. But this assessment ignores human errors. By increasing the number of tasks, you provide more opportunities for respondent errors to cancel themselves out. Increasing the number of tasks in this case (beyond the saturated orthogonal plan) will slightly reduce the reported D-efficiency (from 100% to something slightly less than 100%), but the precision of the estimated parameters may be significantly improved due to a greater amount of information provided by each respondent.

To summarize our point, don’t focus solely on D-efficiency. Good CVA designs foremost should include enough conjoint questions relative to the number of parameters to yield relatively precise estimates of part worths. Given an adequate number of conjoint questions, we next focus on selecting a design with high D-efficiency.

Discarding "Obvious" Tasks

CVA can use *a priori* designations you provide when specifying attributes to identify and omit conjoint profiles that result in obvious answers. Questions that include a product with many good features at a low price or many bad features at a high price are less efficient for deriving conjoint utilities. More balanced questions are more challenging for respondents and can provide more useful information. Moderately balancing the expected utility of the concepts in pairwise questionnaires usually results in better conjoint designs. It may also be beneficial for traditional “card-sort” exercises. However, too much utility balance can lead to difficult questionnaires and increased response error. CVA takes a conservative approach of

only discarding the most “obvious” conjoint tasks. For the advanced user, even more balance can be accomplished by specifying additional prohibitions.

In summary, the CVA designer creates near-orthogonal designs for either single-concept or pairwise designs subject to considerations practitioners commonly face when designing real-world conjoint studies:

- the number of questions (stimuli) to include,
- attribute level prohibitions, and
- whether to discard “obvious” tasks (stimuli loaded with an unrealistic number of good or bad levels).

The designer makes it possible for those who are not specialists in statistics to easily create efficient designs and achieve reliable results.

Part Worth Utility Estimation

With CVA you can use either ordinary least squares (OLS) or monotone regression to calculate utilities. After pretesting your questionnaire it may be helpful to calculate utilities for a sample of respondents using both methods. A comparison of the two kinds of utilities can help you decide which method to use. A unique set of part worths is automatically estimated for each individual, and the utility run is saved for use within the Market Simulator. You can also export the utilities to a text-only data file, an SPSSTM .SAV file or an ExcelTM-friendly .CSV file.

Ordinary Least Squares and Monotone Regression

OLS is the method of calculation traditionally used in most conjoint studies. It can provide valuable diagnostic information about the quality of the calculated utilities. However, OLS is not appropriate for conjoint data consisting of rank orders.

For OLS to be appropriate, we must assume the data are “scaled at the interval level.” By this, we mean that the data are scaled so that real differences in the things being measured are communicated by the arithmetic differences in their values. Fahrenheit temperature, for instance, has an interval scale. The difference between 70 and 80 degrees is exactly as large as the difference between 80 and 90 degrees. In the social sciences and in marketing research we are usually willing to assume that rating scale values possess this kind of scaling.

However, it is usually not reasonable to make such an assumption about rank order data. Suppose a respondent were to rank 30 concept statements in terms of his likelihood of buying each concept. In the absence of other information, we would probably expect the concepts to have a normal distribution of buying likelihood. If so, then we would expect there to be larger “real” differences between concepts with extreme ranks (such as 1 versus 2, or 29 versus 30) than those in the center of the distribution (such as ranks 14 and 15).

When the data are rank orders, it is more appropriate to use a method of calculation that does not assume that the data represent anything more than rank orders. That is the case with nonmetric methods, and in particular with the monotone regression method provided in CVA.

There is also another reason why we have provided a nonmetric method of calculation within CVA: With such methods it is easier to constrain calculated utilities to conform to the researcher’s expectations.

Conjoint utilities are often observed to violate principles of common sense. For example, in pricing studies it sometimes turns out that respondents seem eager to pay higher prices rather than lower prices. This may accurately reflect some respondents’ behavior; price is sometimes taken as an indicator of product quality, and respondents may suspect that a low price reflects poor quality.

However, it is customary to explain to the respondent that “everything else is equal,” and that the attributes are to be considered independently of one another. Under those conditions, if a high price level receives a higher utility than a low price level, we are likely to conclude that the respondent was simply confused. Rather than discard data for that respondent, it is often useful to provide additional information to the calculating program in the form of “constraints.” For example, we may tell the calculating program that the utility for \$1.00 must be no lower than the utility for \$1.25.

With nonmetric methods it is easy to enforce such constraints. Since these methods are iterative, all that is necessary is to insure that the successive estimates at each stage obey the specified inequalities.

With OLS it is much more difficult to enforce such constraints. CVA provides that capability in a limited way: after the least squares solution is computed, then it is adjusted by “tying” values that violate specified inequalities. However, this is an inelegant way of solving the problem. When the data contain many violations of common sense relationships, then the nonmetric method provides a better way of enforcing desired constraints. However, more recent research has shown that HB estimation provides yet an even better way for estimating utilities that are less prone to reversals and for additionally enforcing utility constraints.

Simulating Respondent Preferences

CVA is a component within the SMRT software system, which includes a simple tabulation program and a more powerful Market Simulator. The Market Simulator lets the researcher model a hypothetical “market” by specifying each product's level on each attribute. The file of respondent utility values is read, and a computation is made of each respondent's relative utility for each hypothetical product. There are five options for the way the utilities are used:

1. *First Choice*: Each respondent is allocated to the product having highest overall utility.
2. *Share of Preference*: Each respondent's “share of preference” is estimated for each product. The simulator sums the utilities for each product and then takes antilogs to obtain relative probabilities. Those are then percentaged for each respondent to obtain shares of preference. Shares of preference are averaged for all respondents, and those averages are used to summarize preferences for the respondents being analyzed.
3. *Share of Preference with Correction for Product Similarity*: The “share of preference” model is widely used, perhaps because of its simplicity and because many regard it as an intuitively reasonable model of reality. However, it has a widely recognized disadvantage. Unlike the First Choice model, if an identical product is entered twice, the share of preference model may give it as much as twice the original share of preference. This problem is sometimes referred to as the “red bus/blue bus” or “IIA” (independence from irrelevant alternatives) problem. Our third option was an earlier attempt to overcome this problem. It examines the similarity of each pair of products and deflates shares of preference for products in proportion to their similarities to others. This ensures that the share of two identical but otherwise unique products together will equal what either product alone would get. This technique is available in the software for historical purposes, as a newer technique seems to generally perform better. We generally recommend using the newer fifth option for correcting for product similarity, called Randomized First Choice.
4. *Likelihood of Purchase*: The first three options all assume a set of competitive products, and are concerned with estimating the share a product might receive in that competitive context. However, sometimes no competitive products are available. For example, the researcher might be concerned with a new product category, and might want to estimate absolute level of interest in the category rather than share of preference within the category. Our fourth option does this. The utilities are scaled so that an inverse logit transform provides estimates of purchase likelihood, as expressed by the respondent in the calibration section of the questionnaire. This method is

appropriate if single-concept questionnaires are used and respondents rate the cards on a probability of purchase scale.

5. Randomized First Choice: The Randomized First Choice (RFC) method combines many of the desirable elements of the First Choice and Share of Preference models. As the name implies, the method is based on the First Choice rule, and significantly reduces IIA difficulties. Rather than use the utilities as point estimates of preference, RFC recognizes that there is some degree of error or variance around these points. The RFC model adds unique random variance to each part-worth (and/or product utility) and computes shares of preference in the same manner as the First Choice method. Each respondent is sampled many times to stabilize the share estimates. The RFC model results in a correction for product similarity due to correlated sums of variance among products defined on many of the same attributes.

The CVA Market Simulator provides these capabilities:

- The researcher provides a “base case” containing initial product specifications.
- Each product's base share of preference is first evaluated, and then product specifications can be modified and simulations can be done to estimate the impact of specific modifications.
- The effects of all possible one-attribute-at-a-time modifications can be evaluated automatically.
- The simulator can interpolate between attribute levels.
- Subsets of respondents can be analyzed, and average utilities as well as shares of preference are computed for each subset. Standard errors are also reported.
- Respondents can be weighted using demographic or other data.
- The results can be directly saved to Excel™ files, or cut-and-pasted into most any Windows-based program.

With CVA, we recommend that the researcher include holdout choice tasks in the questionnaire so that the simulator can be tuned (using the “exponent” or scale factor) so that the shares of preference are predictive of choice probabilities.

Deciding Which Conjoint Method to Use

Many methods are available for collecting and analyzing conjoint data, and the researcher contemplating a conjoint study must choose among them. We at Sawtooth Software have had many years of direct experience with these methods, as well as the benefit of many conversations with users of our own and other software. Based on that experience, we offer the following suggestions:

The Full Profile Method (such as used in CVA) was the original conjoint method introduced to the marketing research community, and it remains a standard. Green and Srinivasan (1990) recommend use of the full profile method when the number of attributes is six or fewer. Although the answer depends on the complexity of the attribute descriptions and respondent interest/knowledge of the product category, we agree that six is an appropriate cutoff number. We think respondents are likely to become overloaded and confused when confronted by large numbers of lengthy profiles. Our experience is that, when there are more than about six attributes, and pricing research is not the goal, ACA works better--particularly if hierarchical Bayes estimation is employed. We also think the weight of evidence shows that ACA works at least as well as full profile when there are fewer than six attributes (for example, see Huber et al., 1993) and pricing research is not the goal, although with few attributes ACA has no compelling advantage.

The ACA System was developed specifically for the situation where there are many attributes and levels. Most of ACA's questions present only small subsets of attributes, so questions do not necessarily become more complex when there are many attributes in the study. With more than six attributes, we think ACA is likely to be the more appropriate method when pricing research isn't the goal. ACA can capture a great deal of information from respondents in a relatively short amount of time, and therefore has advantages for dealing with many attributes and when using relatively small sample sizes.

Like most full profile applications, ACA is a "main effects only" model, and assumes there are no interactions among attributes. Many conjoint practitioners agree that one must remain alert for the possibility of interactions, but that it is usually possible to choose attributes so that interactions will not present severe problems. ACA has been shown to have weaknesses in pricing research, where it often underestimates the importance of price. We generally recommend that either CVA or CBC be used if pricing research is the main purpose of your study.

The CVA System is conjoint software first introduced by Sawtooth Software in 1990 for traditional full-profile conjoint analysis. It is a good technique when the number of attributes is about six or fewer, and also when dealing with relatively small sample sizes. It often does a better job than ACA in pricing research. CVA uses either a paired-comparison or a single-concept interview that can be administered by computer or with paper and pencil.

The CBC System is conjoint software first introduced by Sawtooth Software in 1993 to administer and analyze "Choice-Based Conjoint" studies. CBC conducts a paper- or computer-administered interview in which the respondent sees a series of choice tasks. Each task displays several concepts and asks which the respondent would choose from that set. Optionally, a "would choose none" option may be offered. Attribute levels in each concept are varied in such a way that values similar to conjoint utilities can be estimated for each attribute level. Analysis can be done at the group level with multinomial logit analysis, which is included with the base CBC system. Additionally, latent segment-based utilities can be generated using Latent Class. Individual-level utilities can be estimated from choice data using hierarchical Bayes (CBC/HB).

We think CBC provides three potential advantages over other conjoint methods:

1. It presents tasks that may be more "realistic" than other conjoint methods. In the real world, buyers express their preferences by choosing one product or another, rather than by rating or ranking them.
2. By including the opportunity for the respondent to choose "None of these," CBC may be able to deal more directly with questions relating to volume (rather than just share). By contrast, ACA models volume using "Likelihood of Purchase" simulations, based on responses to Calibrating Concepts.
3. Because CBC analysis can be done for groups rather than for individual respondents, sufficient information is available to measure interactions as well as main effects.

However, CBC has the disadvantage of being an inefficient way of collecting data. The respondent must read and process several full profile concepts before giving each answer. To keep the respondent from becoming overloaded and confused, we suggest using no more than about six attributes with CBC. CBC should be considered when there are few attributes and when interactions are likely to occur, both of which are often true of pricing studies. CBC studies also require a bit larger sample size than ACA and CVA to achieve equal precision of estimates.

One of the greatest advantages for using CBC is that the resulting utilities are scaled based on the choice data. When applied in market simulations, these choice-developed utilities lead to shares of preference for competitive products that are directly linked to probabilities of choice. In contrast, CVA utilities are estimated from rating scales or rank-order data. These scales do not automatically lead to market simulations with appropriate choice probability scaling.

CVA users can achieve approximate choice-probability scaling for utilities by adding holdout choice tasks within their CVA surveys. By tuning the scale factor (exponent) within Share of Preference or Randomized First Choice simulations to best fit holdout choice probabilities, the simulator may be tuned to resemble choice-based scaling.

CVA System Specifications

Questionnaire Design

Questionnaire format: Paired comparison or Card-sort in full-profile.

Maximum number of attributes:	30
Maximum number of levels per attribute:	15
Maximum number of questions	500

Utility Calculation

Choice of algorithm: Least Squares or Monotone Regression (an add-on module is available for hierarchical Bayes (HB) estimation).

Can accept data directly from text-only (*.txt) files

Market Simulator

Choice Models: First Choice
Share of Preference
Share of Preference with Correction for Product Similarity
Purchase Likelihood
Randomized First Choice

Appendix

How CVA Calculates Utilities

Ordinary Least Squares (OLS)

This section assumes the reader has basic knowledge of linear regression. However, understanding of the technical details in this section is not essential for using CVA.

The OLS calculator does ordinary least squares linear regression using a “dummy variable” approach. A vector of independent variables for each observation (conjoint question) is built using information from the experimental design. The vector has elements of 1, 0, or -1, depending on whether respective attribute levels appear in that question, and whether they appear on the left-hand or right-hand side of the pairwise questions.

The dependent variable for each observation is obtained by applying the indicated recoding transformation to the corresponding data value. If the data value has the code reserved for missing data, then that observation (conjoint question) is not included in the calculation.

When using regression to estimate conjoint utilities, it is customary to delete one level of each attribute from the computation. Otherwise, there is a linear dependence among the variables describing levels of each attribute, which leads to indeterminacy in the computation. Omitting one level of each attribute from the computation is equivalent to setting its utility at zero, with the other levels measured as contrasts with respect to zero. The OLS calculator omits the first level of each attribute from the regression computation. Thus, if there are k attributes with a total of n levels, the regression is done with only $n-k$ independent variables. The indeterminacy could also be handled by adding side conditions, such as requiring that the utilities for each attribute sum to some constant. However, our approach has the advantage of greater computational speed.

An intercept term is also computed by CVA, but it is not reported separately. Since the utilities will most often be used by adding up sums consisting of one value from each attribute, the intercept has been divided by the number of attributes and that fraction has been added to every utility value. Thus the first level for each attribute, which would otherwise be zero, will be equal to the intercept divided by the number of attributes.

The “ r squared” value for each respondent does not contain any correction for degrees of freedom. If the number of observations is equal to the number of parameters being estimated (levels - attributes + 1), then the r squared value will be unity.

If the design is deficient – containing either too few observations to permit estimation or insufficient information for a particular attribute level – then a message to that effect will appear on the screen and utilities will not be estimated for that respondent.

If there are degrees of freedom available for error, then descriptive data will be written with information about the precision of estimation.

A statistic (“rms cor”) is provided for each respondent, which describes the amount of correlation among the independent variables. It is the “root mean square” of off-diagonal elements of the correlation matrix for the $n-k$ independent variables. Subtle relationships among variables can easily lead to faulty designs that would not be detected, and therefore we caution against paying much attention to this statistic. We include it only because our users may be accustomed to similar statistics in other software packages.

In an orthogonal design with no missing data, this value will be either zero or a small positive number. (Orthogonal designs have correlations within attributes.) Orthogonal designs are sometimes altered to eliminate “nonsense” questions, which can compromise orthogonality. Also, some design procedures

(CVA's questionnaire design module, for example) produce well-balanced but not perfectly orthogonal designs.

For each individual, standard errors of utility estimates are provided, except for the first level of each attribute, which is assumed to have utility of zero. These standard errors may also be regarded as standard errors of differences between each level and the first level of that same attribute. These standard errors can be of diagnostic value. Attribute levels with large standard errors should be given more attention in questionnaire design. They may appear in too few questions, or they may occur in patterns that compromise the level of independence necessary for good estimation.

Another statistic ("rel effec") describes the quality of the conjoint design. We are mainly interested in the precision of estimates of differences in utility among levels of the same attribute. The relative error variance of the difference is obtained from elements of the inverse of the correlation matrix among independent variables. The relative error variance of the difference between two utility estimates is obtained by summing those two diagonal elements of the inverse and subtracting twice the corresponding off-diagonal element. We compute the relative error variances of all within-attribute pairwise differences.

Suppose there were an orthogonal design for the same attributes and levels, and all the levels of a particular attribute appeared the same number of times. (Such a design may or may not actually exist.) Then we could compute similar relative error variances for that "ideal" design.

Our "Relative Efficiency" index is the ratio of the sum of theoretical relative error variances for such an ideal design, divided by corresponding sum of relative error variances for the actual design used. The overall relative efficiency index is reported, as is an index for each attribute. The best possible value is 1.000, although values that high may not actually be achievable for some combinations of numbers of attributes and levels, since the required orthogonal designs may not exist.

These coefficients can be useful diagnostics for a conjoint questionnaire design during the pilot stages of a study. If the coefficient is too low for a particular attribute, then additional questions should be introduced, or the design should be modified to have less dependence among the attribute levels.

Monotone (nonmetric) Regression

This option for calculating utilities uses a method described by Richard M. Johnson in "A Simple Method of Pairwise Monotone Regression", *Psychometrika*, 1975, pp 163-168.

The method is iterative, finding successive solutions for utility values that fit the data increasingly well. An initial solution is developed, either randomly or using information in the experimental design. Two measures of goodness of fit are reported: theta and tau.

Tau

Suppose the conjoint questionnaire presented concepts one at a time and asked for a rank order of preference. Although there would have been many concepts in the questionnaire, consider just four of them, concepts P, Q, R, and S. Suppose the respondent ranked these concepts 7, 9, 13, and 17, respectively, and at some intermediate stage in the computation, utilities for these concepts are estimated as follows:

Concept	Estimated Utility	Preference Rank
P	4.5	7
Q	5.6	9
R	1.2	13
S	-2.3	17

We want to measure "how close" the utilities are to the rank orders of preference.

One way we could measure would be to consider all of the possible pairs of concepts, and to ask for each pair whether the member with the more favorable rank also has the higher utility. Since these are rank orders of preference, smaller ranks indicate preference, so we know that:

Preference	Utility Difference	Squared Difference
P is preferred to Q	-1.1	1.21
P is preferred to R	3.3	10.89
P is preferred to S	6.8	46.24
Q is preferred to R	4.4	19.36
Q is preferred to S	7.9	62.41
R is preferred to S	3.5	12.25
Total		152.36

Of the six pairs, five have utility differences with the correct signs (the preferred product has the higher utility), and one pair has a utility difference with the wrong sign.

Kendall's tau is a way of expressing the amount of agreement between the preferences and the estimated utilities. It is obtained by subtracting the number of "wrong" pairs from the number of "right" pairs, and then dividing this difference by the total number of pairs. In this case,

$$\text{tau} = (5 - 1) / 6 = .667$$

A tau value of 1.000 would indicate perfect agreement in a rank order sense. A tau of 0 would indicate complete lack of correspondence, and a tau of -1.000 would indicate a perfect reverse relationship.

Tau is a convenient way to express the amount of agreement between a set of rank orders and other numbers, such as utilities for concepts. However, it is not very useful as a measure on which to base an optimization algorithm. As a solution is modified to fit increasingly well, its tau value will remain constant and then suddenly jump to a higher value. Some other measure is required that is a continuous function of the utility values.

Theta

For this purpose we use the statistic "theta." Theta is obtained from the squared utility differences in the last column of the table above. We sum the squares of those utility differences that are in the "wrong order," divide by the total of all the squared utility differences, and then take the square root of the quotient. Since there is only one difference in the wrong direction,

$$\text{theta} = \text{square root}(1.21/152.36) = .089$$

Theta can be regarded as the percentage of information in the utility differences that is incorrect, given the data. The best possible value of theta is zero, and the worst possible value is 1.000.

Now that we have defined theta, we can describe the nature of the computation.

An initial solution is obtained, either using random numbers or the information in the experimental design.

With each iteration a direction is found which is most likely to yield an improvement. A "line search" is made in that direction, and the best point on that line is determined. Each iteration has these steps:

1. Obtain a current value of theta and a direction (gradient vector) in which the solution should be modified to decrease theta most rapidly.

2. Take a step in that direction and recompute theta. If theta is larger than before, try a smaller step. Continue trying smaller steps until a smaller value of theta is found. (If no smaller value of theta can be found in this way, the process has converged and the computation ends.)
3. After a smaller value of theta has been found, continue taking steps of the same size until a value of theta is found that is larger than the previous one. This means that we have gone too far.
4. Fit a quadratic curve to the last three values of theta to estimate the position of the minimal value of theta along this line.
5. Evaluate theta at the estimated optimal position. If it is the best value of theta found so far, end the iteration with that solution. Otherwise end the iteration with the best solution found.
6. Adjust the step size by dividing by the number of successful steps in this iteration. (The goal is to have exactly one successful and one unsuccessful step.)

A record of the computation is presented on the screen. Here is the screen display for our sample problem:

Computing for respondent number: 123456

```
Iteration 1: theta= 0.1679 tau= 0.7541 steps= 1
Iteration 2: theta= 0.1050 tau= 0.8142 steps= 1
Iteration 3: theta= 0.0896 tau= 0.8142 steps= 3
Iteration 4: theta= 0.0885 tau= 0.8033 steps= 1
Iteration 5: theta= 0.0860 tau= 0.8087 steps= 3
```

In theory, the iterations could continue almost indefinitely with a long series of very small improvements in theta. For this reason it is useful to place a limit on the number of iterations.

CVA has a default limit of ten iterations. The actual number of iterations may be less than this, because the computation may terminate earlier as in point 2 above. Each successive estimate of utilities is constrained as indicated by the *a priori* settings or additional utility constraints.

As with many nonmetric procedures, there is no guarantee that the solution found is the “global minimum.” That is to say, there may be other possible solutions that would be as good or better. However, experience in dozens of applications has indicated that this is not a problem. Computations starting from different initial solutions appear to converge to similar solutions, and in actual practice there does not seem to be a problem with “local optima.”

How CVA Utilities Are Scaled

Monotone regression

CVA's monotone regression utility calculator scales utilities in a way that is easy to describe and to understand. For each respondent, the values for each attribute have a mean of zero, and their sum of squares across all attributes is unity. Here is an example, assuming two attributes, one with 3 levels and one with 2 levels:

	utility	square
Attribute One Level 1	.50	.25
Level 2	.10	.01
Level 3	-.60	.36
Total	0.00	
Attribute Two Level 1	.44	.19
Level 2	-.44	.19
Total	0.00	1.00

OLS regression

CVA's OLS utility calculator scales utilities in a way that depends upon the data, and upon the researcher's use of the recode capabilities. The calculation has these steps:

1. If automatic recoding was specified, then the data are automatically recoded. If no recode was specified, the values in the data file are used without modification.
2. An array of "dummy" variables is constructed with a row for each conjoint question and a column for each attribute level. Each cell of this array has a value of 1, 0, or -1, depending on the experimental design. For single-concept presentation, the values are either 1 or 0. If the level appears in the concept it is coded as 1, and if absent it is coded as 0. For pairwise presentation, if a level appears in the left-hand concept it is coded as -1, or 1 if in the right-hand concept. If an attribute level does not appear in either profile, then the corresponding array element is 0.
3. The first level (column) for each attribute is omitted temporarily from the design, which avoids technical problems of indeterminacy in the solution. (See Appendix C for more information about linear dependency.)
4. OLS regression is used to predict the transformed data values from the surviving columns of the array (variables). A regression coefficient is computed for each variable, as well as a single intercept. The regression coefficients for the omitted variables are assumed to be zero.
5. The intercept is divided by the number of attributes, and the quotient is added to every regression coefficient, including those previously assumed to be zero. The resulting values are reported as utilities for the attribute levels. (The intercept is handled in this way to make it easy to calculate total utilities for products during simulations. Since each product to be simulated will have exactly one level from each attribute, the simulator will be able to include the intercept automatically just by adding the utilities of its attribute levels.)

As can be seen from this explanation, with the OLS calculator the scaling of the utilities is completely under the control of the researcher. Other things being equal, if the data are collected on a 100-point scale, the utilities will be about ten times the magnitude as they would be if the data were collected on a 10-point scale.

References

Green, Paul E. and V. Srinivasan (1990), "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," *Journal of Marketing*.

Huber, Joel, Dick Wittink, John Fiedler, and Richard Miller (1993), "The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice," *Journal of Marketing Research*, 30 (February), 105-114.