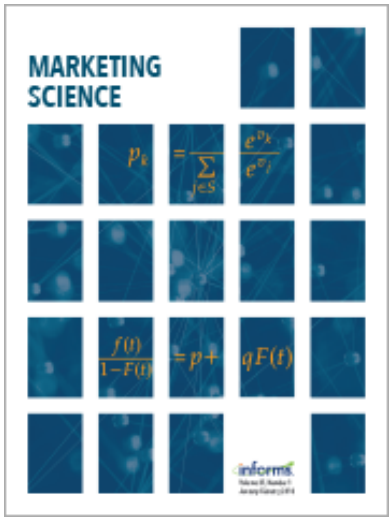


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

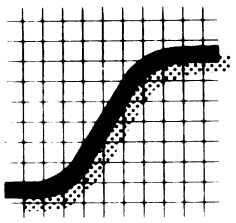
Technical Note—Factors Influencing the Selection of Preference Model Form for Continuous Utility Functions in Conjoint Analysis

Philippe Cattin, Girish Punj,

To cite this article:
Philippe Cattin, Girish Punj, (1984) Technical Note—Factors Influencing the Selection of Preference Model Form for Continuous Utility Functions in Conjoint Analysis. Marketing Science 3(1):73-82. https://doi.org/10.1287/mksc.3.1.73

MARKETING SCIENCE
Vol. 3, No. 1, Winter 1984
Printed in U.S.A.

TECHNICAL
Note



FACTORS INFLUENCING THE SELECTION OF PREFERENCE MODEL FORM FOR CONTINUOUS UTILITY FUNCTIONS IN CONJOINT ANALYSIS

PHILIPPE CATTIN AND GIRISH PUNJ
University of Connecticut

In conjoint analysis, a consumer's utility function for a continuous attribute is usually estimated using a part worth function. However, one may also use continuous functions. The purpose of the paper is to investigate through simulation the combined influence of the number of degrees of freedom, the nature of the "true" utility function and the amount of error in the data on the selection of a utility function. The focus is on functions that are known or expected to be monotone within the range of attribute levels of interest. One can then choose among linear, quadratic and part worth functions. The results show (a) that estimation procedures with monotonicity constraints should be used, and (b) that it is best to use quadratic or part worth functions rather than linear functions to minimize potential losses in predictive validity.

(Preference Model; Conjoint Analysis; Constrained Utility Functions)

Introduction

In conjoint analysis a consumer's utility for a continuous attribute is often estimated for several discrete levels of the attribute using a part-worth function model. The utility of an intermediate level is then interpolated, if needed. Alternatively, a continuous utility function can be assumed and estimated. While only the part worth function model can be used with categorical attributes, continuous attributes can be represented with several models. There are three major types of models: the vector model, usually represented by a linear function, the ideal point model, usually represented by a quadratic function, and the part worth function model. Flexibility increases from the vector, to the ideal point, and to the part worth function models allowing more shapes for a utility function. However, the number of parameters estimated increases as one goes from the vector to the part worth model (if there are more than three levels), making the parameter estimates for the part worth model least reliable (Green and Srinivasan 1978, p. 106).

Other considerations affecting the choice of a model are the nature of the "true"

0732-2399/84/0301/0073$01.25

preference model underlying the data (which is seldom known in practice) and the amount of error (noise) in the data. The purpose of this study is to investigate the combined influence of the number of degrees of freedom, the nature of the "true" preference model and the amount of error in the data on the selection of a preference model. To our knowledge, there has been no systematic effort to investigate the combined influence of these three factors.

Previous Research and Focus of This Study

Most attribute utility functions used in conjoint analysis fall into one of the four cases shown in Figure 1.[1] The quadratic and part worth functions can be used in all four cases, while the linear function can be used in three out of four cases. Other

Expectation Concerning[b] Utility of Attribute i	Example	Appropriate Mathematical Function[a] over the Range of Levels of Interest	Constraint[c] on the Parameters (within the Range of Interest)
1. Ideal Point	Sugar Content (over a wide range)	Quadratic	$b_i \leqslant 0$
2. Monotone Increasing	Miles per Gallon	(2a) Linear	$a_i \geqslant 0$
		(2b) Quadratic	$a_i + 2b_iX_i \geqslant 0$
3. Monotone Decreasing	Price	(3a) Linear	$a_i \leqslant 0$
		(3b) Quadratic	$a_i + 2b_iX_i \leqslant 0$
4. Ideal Point or Monotone (decreasing or increasing)	Sugar Content	(4a) Linear	(no constraint)
		(4b) Quadratic	$b_i \leqslant 0$ at $a_i + 2b_iX_i = 0$ if $a_i + 2b_iX_i = 0$ within the range of interest

[a] The part worth function is not shown, but can be used in all cases.

[b] The four expectations included in this figure, and their appropriate mathematical representation do not include multiple peaks. In this case, not only can the part worth function be used, but also a cubic or higher order function. However, as argued in the paper, such cases do not seem frequent, at least at the individual level.

[c] The constraints on the parameters were obtained by taking the derivatives of the functions (e.g., the derivative of $a_iX_i + b_iX_i^2$ is $a_i + 2b_iX_i$).

FIGURE 1. Major Types of Utility Functions. (The linear function is represented by a_iX_i and the quadratic function by $a_iX_i + b_iX_i^2$ where a_i and b_i are the parameters and X_i the level of the attribute that varies over the range of levels of interest.)

[1] Sometimes, one may expect multiple peaks. Such an expectation is not included in Figure 1. Tea (if the range of levels goes from cold to hot) falls in the multiple peaks category because consumers tend to prefer iced and hot tea to in-between temperature levels (Green and Srinivasan 1978, p. 106). This implies two peaks (ideal points). But then one could argue that iced tea and hot tea belong to two different product categories. Moreover, as argued by Pekelman and Sen (1979a, p. 266), this type of function seems more likely to occur when aggregating responses across consumers than at the individual level. For instance, there are consumers who prefer low suds detergent and others high suds detergent, which gives two ideal points at the aggregate level (Kuehn and Day 1962, Exhibit 2), but not as the individual level. Since the concern here is with individual-level models (which are used in the market simulations often done in conjoint analysis studies), the multiple peaks case is not considered important.

nonlinear monotone continuous functions are available for use as well. However, they are not easily linearizable in the parameters, and are not used in conjoint analysis studies.[2]

This study is confined to attributes with 2 to 4 levels because this is the range of levels most frequently encountered in conjoint analysis studies. *Two-level* attribute models need not be formally investigated because the linear, quadratic and part worth functions estimate the same number of parameters and produce the same attribute utility estimates for the two levels used in the study and all intermediate levels.

When using models with *three-level* attributes, the quadratic and part worth functions estimate two parameters each, excluding the intercept, while the linear function estimates one. The attribute utilities and the predictive validities of the quadratic and part worth models for the three levels used in the estimation are identical. For intermediate levels, attribute utility estimates differ since interpolations are necessary for the part worth function. With *four-level* attribute models the utility estimates obtained with the three functions are in general different for both estimated and intermediate levels.

The present study concentrates on monotone functions, using estimation procedures with monotonicity constraints, because that is where there is a wide choice of models available. In fact, with three level attributes and a monotonicity constraint, the quadratic and part worth models can produce different results for the levels included in the study. Little is known about the impact of the monotonicity constraints on the utility estimates. On the other hand, in the nonmonotone case (or unconstrained monotone) the choice of models is more limited and exists only for the four-level attribute case and beyond.

Previous research (Pekelman and Sen 1979a, b) has shown that the use of quadratic functions improves predictions over part worth functions, when the underlying data are quadratic and the number of attribute levels is four. This finding is expected because the part worth function requires more parameter estimates than the quadratic function (three versus two excluding the intercept). Thus the parameter estimates of the part worth function are less reliable.

Two Monte Carlo simulations were performed to study the combined effect of the amount of noise in the data, of the nature of the true function and of the number of degrees of freedom, measured by the number of observations/number of parameters (N/p) ratio. The first simulation used six three-level attributes, while the second one used four four-level attributes. The resulting number of parameter estimates is representative of real world problems.

Three types of data were produced: linear, quadratic and part worth. The quadratic functions used to produce the data were selected so as to depart as much as possible from linearity (subject to being monotone within the range of attribute levels of interest). Hence, the resulting quadratic data are as favorable as possible to the quadratic function (as opposed to the linear function), just as the linear data are as favorable as possible to the linear function. The part worth data which were produced were (about) as nonlinear as possible, and thus as favorable as possible to the part worth function (as compared to the linear and quadratic functions). It is not claimed

[2] For instance, the logarithmic and the exponential functions are monotone (increasing or decreasing) and nonlinear and could be used for cases 2 and 3 of Figure 1. But then, their mathematical representations (e.g., $U(X_{ij}) = \log(a_i + b_i X_{ij})$ for the logarithmic function and $U(X_{ij}) = \exp(a_i + b_i X_{ij})$ for the exponential function, where a_i and b_i are parameters) are not linearizable in a_i and b_i. If one attribute is represented by a logarithmic or exponential function and another attribute by another function, the resulting multiattribute model is nonlinear and a search procedure is required. On the other hand, the quadratic function $(q_i X_{ij} + r_i X_{ij}^2)$ has one linear term in X_{ij} and the other linear in X_{ij}. Hence, the linear, quadratic and part worth functions are the only functions investigated in this study.

that these three types of data represent real data. Real data are unknown. But then these three types of data were produced by the three models commonly assumed in practice and will enable us to determine how much one loses by using each model (inappropriately) on the other two data types. Moreover, these three types of data are extreme and real data are likely to be "in between."

First Simulation Study

Data Generation

This simulation was conducted using six attributes with three levels for each attribute. Orthogonal arrays of sizes 18, 27 and 36 were used to construct the hypothetical stimuli. Since both the quadratic and part worth functions require two parameter estimates (excluding the intercept) per attribute, the resulting N/p ratios for these models were 1.50, 2.25 and 3.00, respectively. The three levels of each attribute were assigned the values 1, 2 and 3. (The intermediate level was defined as the middle point between the two extremes, because this is typically what is done in practice.)

For each hypothetical sample of stimuli derived from the orthogonal array, the *linear data* were produced using the model

$$U_j = \sum_{i=1}^{6} a_i X_{ij} + \epsilon_j \qquad \text{where} \tag{1}$$

X_{ij} is the level of attribute i of stimulus j,
a_i is the slope parameter corresponding to attribute i,
U_j is the utility of profile j and
ϵ_j is the error.

The a_i slopes were randomly drawn from the positive half of a standard (0, 1) normal distribution. Only positive values of a_i were used because the sign of the a_i's has no influence on the predictive validities. The ϵ_j errors were drawn from a normal distribution with mean zero and variance σ^2, where σ^2 was set such that the expected percentage of noise in the data took on the values 11.5, 26 and 65 percent. These percentages correspond to high, medium and low levels of predictive validity found in real data (e.g., Scott and Wright 1976).[3] The a_i, ϵ_j and X_{ij} values were combined to produce a vector of interval scaled U_j values for each sample of data. The linear data thus produced are as favorable as possible to linear models, when recovering the data.

The *quadratic data* were produced using the model

$$U_j = \sum_{i=1}^{6} a_i \left(3X_{ij} - 0.5X_{ij}^2\right) + \epsilon_j . \tag{2}$$

The a_i and the ϵ_j values were obtained using the same procedures as with the linear data. The coefficients corresponding to the linear and quadratic terms (3 and -0.5 respectively) were selected so that the functions are monotonically increasing between 1 and 3 (the range of levels of interest) and reach their maxima at $X_{ij} = 3$. As a consequence, the resulting quadratic data depart from linearity as much as possible (subject to being monotone in the range of interest), and are as favorable as possible to quadratic models (compared to linear models), when recovering the data.

The *part worth data* were produced in two stages. The utility of attribute i for

[3]The empirical results obtained by Scott and Wright (1976) indicate that high, medium and low levels of predictive validity are 0.90 to 0.95, about 0.8, and about 0.5, respectively. Some preliminary (empirical) work was done to determine what expected percentages of noise in the data would produce such predictive validities: 11.5, 26.0 and 65.0% (respectively) were chosen.

stimulus j was computed using the function:

$$U(X_{ij}) = 0.1X_{ij} \qquad \text{for} \quad 2 \leqslant X_{ij} \leqslant 3, \quad \text{and}$$

$$U(X_{ij}) = 1.9X_{ij} - 3.6 \qquad \text{for} \quad 1 \leqslant X_{ij} < 2.$$

The function is (almost) as nonlinear as possible (subject to being monotone in the range of interest) since the magnitude of the slope changes from 1.9 to 0.1, as X_{ij} goes from the intervals [1, 2] to [2, 3]. This function could represent a respondent who uses a cutoff point for X_{ij} in the interval [1, 2]. The intercept (-3.6) ensures that the difference $U(X_{ij} = 3) - U(X_{ij} = 1)$ is the same for both the quadratic and part worth functions. The part worth data were then produced using the model

$$U_j = \sum_{i=1}^{6} a_i U(X_{ij}) + \epsilon_j \tag{3}$$

where the a_i and ϵ_j values were obtained using the same procedure as with the linear data. The resulting part worth data were (almost) as favorable as possible to part worth models (compared to linear and quadratic models), when recovering the data.

The simulation design consisted of nine cells corresponding to three sample sizes (18, 27 and 36) and three noise levels (low, medium and high). Three types of data (linear, quadratic and part worth) were produced in each cell. The number of replications per cell and per data type was set at 20 to achieve a balance between computation cost and estimation accuracy. Twenty replications were also used in the simulations reported by Srinivasan (1975) and Pekelman and Sen (1979a).

For each replication, a validation sample of size 729 was generated using all combinations (3^6) of attribute levels for the six attributes. The utility values (U_j's) for the validation data were obtained using equations (1), (2) or (3), except that no errors (ϵ_j's) were included in the computations. In this manner, it takes fewer replications per cell to obtain significant differences. The resulting measures of predictive accuracy overestimate the predictive accuracy obtained with real data. Since the purpose of the simulation was to *compare* predictive accuracies (as opposed to obtaining accurate estimates), the above fact was not viewed as a limitation.[4]

Estimation

OLS regression was used on the three types of data to estimate a linear, a quadratic and a part worth model for each replication in each cell of the design (i.e. $20 \times 9 = 180$ linear, quadratic and part worth models were estimated). The reason for using OLS regression was convenience and lower computing costs. The predictive validity obtained with OLS regression has been found to be about as good as that obtained with LINMAP or MONANOVA (Green and Srinivasan 1978, p. 114). Also, regression is now used more often than LINMAP or MONANOVA (Cattin and Wittink 1982, Table 7). All the $180 \times 3 = 540$ models were also estimated with a monotonicity constraint on the estimated functions. The constraints applied are discussed below.

[4]Validation data corresponding to intermediate values of X_{ij} (i.e., between 1 and 2, and 2 and 3) were not generated. Previous research (Cattin 1982, footnote 6) has shown that the relative performance of linear, quadratic and part worth models does not change appreciably with the use of intermediate values of the X_{ij}'s. Validation data corresponding to values of X_{ij} outside the range were not generated either. The main reason is that the range should be defined so as to include the real and realistic levels (Green and Srinivasan 1978, p. 109). If one were to estimate attribute utilities outside the range, one would have to extrapolate the function in such a way that it remains monotone. Hence, a quadratic utility function should be extrapolated until it reaches its maximum. The maximum utility value would then be used for X_{ij} values beyond the X_{ij} value that corresponds to the maximum.

For the quadratic model, a constrained least squares procedure derived from Cunia (1964, p. 565) was used. Cunia's procedure applies to one quadratic attribute utility function at a time. But since the estimation data used in this study are derived from orthogonal arrays, one may use Cunia's procedure on each function independently of the other five functions. For each OLS-estimated quadratic function of the form

$$U(X_{ij}) = q_i X_{ij} + r_i X_{ij}^2,$$

the constraints were applied as follows:

(1) Concave functions with a maximum for $X_{ij} < 3$ were constrained to have their maximum at $X_{ij} = 3$ if $U(X_{ij} = 3) > U(X_{ij} = 1)$, otherwise $U(X_{ij})$ was set equal to zero (i.e., $q_i = r_i = 0$).

(2) Convex functions with a minimum for $X_{ij} > 1$ were constrained to have their minimum at $X_{ij} = 1$ if $U(X_{ij} = 3) > U(X_{ij} = 1)$, otherwise $U(X_{ij})$ was set equal to zero.

Each OLS-estimated linear function was constrained to be monotonically increasing. When this condition was violated $U(X_{ij})$ was set equal to zero. For each OLS-estimated part worth function the constraints were applied as follows:

(1) When $U(X_{ij} = 3)$ was found to be lower than $U(X_{ij} = 2)$ the two estimated part worth utilities were set equal to the average of the two U's.

(2) When $U(X_{ij} = 2)$ was found to be lower than $U(X_{ij} = 1)$, the two estimated part worth utilities were set equal to the average of the two U's. If $U(X_{ij} = 3)$ was also found to be lower than $U(X_{ij} = 1)$, the three estimated part worth utilities were set equal to the same value (e.g., zero).

All the estimated nonconstrained and constrained models were then used to estimate the utility of the stimuli in the validation samples. Pearson correlations were computed between actual and estimated utilities.

Results

The mean correlations between the actual and estimated utilities were computed for the linear, quadratic and part worth models, for each cell of the simulation design. They are shown in Table 1. The correlations obtained with the constrained estimation procedures (shown in parentheses) are substantially superior to those obtained in the nonconstrained case (without parentheses). This is particularly true for the quadratic and part worth models, for high noise levels and small sample sizes. Thus, when a function is known or expected to be monotone, it is not only appropriate, but best, to

TABLE 1

Mean Pearson Correlations Obtained in the Three-Level Attributes Case[a]

			Noise Level		
			High	Medium	Low
Sample Size	High	L	0.874 (0.885)	0.931 (0.932)	0.942 (0.942)
		Q	0.799 (0.893)	0.949 (0.963)	0.980 (0.981)
		PW	0.799 (0.874)	0.949 (0.961)	0.980 (0.984)
	Medium	L	0.790 (0.824)	0.914 (0.919)	0.936 (0.937)
		Q	0.733 (0.837)	0.934 (0.951)	0.974 (0.976)
		PW	0.733 (0.823)	0.934 (0.951)	0.974 (0.979)
	Low	L	0.706 (0.799)	0.895 (0.906)	0.929 (0.932)
		Q	0.645 (0.810)	0.897 (0.928)	0.959 (0.967)
		PW	0.645 (0.793)	0.897 (0.924)	0.959 (0.969)

[a] L stands for linear, Q for quadratic and PW for part worth. Throughout the table, the entries in parentheses correspond to the results obtained with the constrained estimation procedures, while those without parentheses stand for the results obtained with OLS.

use a constrained estimation procedure. Hence, the discussion in the remainder of this section will focus only on the results obtained with the constrained estimation procedures.

The part worth model outperforms the other models when the noise level is low, while the quadratic model is best for high or medium noise levels, whatever the sample size. The improvement in predictive validity achieved with the quadratic model (over the part worth model) goes up to 0.019 (0.893 − 0.874) in the high noise level-high sample size cell. The corresponding figure for the part worth model is quite small (0.003) for the low noise level cells. The predictive validities in Table 1 are averages obtained across replications and data types. Further insight can be gained by disaggregating the results across data types.[5] Such results show that the average loss suffered when using the part worth model goes up to 0.060 with linear data and 0.025 with quadratic data. When using the quadratic model, it goes up to 0.016 at the most with part worth data, and 0.024 with linear data. When using the linear model it goes up to 0.107 with part worth data. Thus the implication is that one might use the quadratic model all the time. Nevertheless, one would not lose much with the part worth model.

Second Simulation Study

Data Generation

This simulation was conducted using four attributes, four levels for each attribute. Orthogonal arrays of sizes 12, 32 and 48 were used to construct the hypothetical stimuli.[6] Since the part worth function requires three parameter estimates per attribute (excluding the intercept), the resulting N/p ratios were 1.33, 2.66 and 4.0. The four levels of each attribute were assigned the values 1, 2, 3 and 4.

For each hypothetical sample of stimuli derived from an orthogonal array, *linear data* were produced using the model

$$U_j = \sum_{i=1}^{4} a_i X_{ij} + \epsilon_j . \tag{4}$$

The a_i and ϵ_j values were drawn as for the linear data in the first simulation study.

The *quadratic data* were produced using the model

$$U_j = \sum_{i=1}^{4} a_i \left(4X_{ij} - 0.5X_{ij}^2\right) + \epsilon_j . \tag{5}$$

The a_i and ϵ_j values were drawn as before. The coefficients corresponding to the linear and quadratic terms (4 and -0.5 respectively) were selected so that the functions are monotonically increasing between 1 and 4 and reach their maxima at $X_{ij} = 4$.

The *part worth data* were produced in two stages, as before. The attribute utilities, $U(X_{ij})$'s, were computed using the function:

$$U(X_{ij}) = 0.1X_{ij} \qquad \text{for} \quad 2 \leqslant X_{ij} \leqslant 4, \quad \text{and}$$

$$U(X_{ij}) = 4.3X_{ij} - 8.4 \qquad \text{for} \quad 1 \leqslant X_{ij} < 2.$$

[5] These disaggregated results are in (Cattin and Punj 1983) which can be obtained from either author. Table 3 of this working paper shows the average loss in predictive validity resulting from using the linear, quadratic and part worth models when the underlying data correspond to a different model form (e.g., using the quadratic model to estimate part worth data), for each cell of the design. There are 54 average losses because there are six model comparisons for each of the nine cells.

[6] A sample size of 48 is a little higher than in typical studies. However, there is no smaller size of orthogonal array to choose from except for 16, 32 and 48. In the end, it does not matter much at all because the relative results of each model are not a function of sample size.

This function could represent a respondent who uses a cutoff point for X_{ij} in the interval [1, 2]. The rationale for using the coefficients shown is the same as for the part worth functions used in the first simulation study. The part worth data were then produced using the model

$$U_j = \sum_{i=1}^{4} a_i U(X_{ij}) + \epsilon_j, \tag{6}$$

where the a_i and ϵ_j values were obtained as before.

As earlier, the simulation design consisted of nine cells, corresponding to three sample sizes (16, 32 and 48) and three noise levels (low, medium and high). Three types of data (linear, quadratic and part worth) were produced in each cell, and there were 20 replications per cell. For each replication, a validation sample of size 256 was generated using all combinations (4^4) of attribute levels for the four attributes. The utility values for the validation data did not include any noise, for reasons mentioned earlier.

Estimation

OLS regression and constrained estimation procedures (described in the first simulation study) were used to estimate linear, quadratic and part worth models for each replication and each cell of the design (i.e., $20 \times 9 = 180$ linear, quadratic and part worth models were estimated). The estimated models were used to estimate the utilities of the profiles in the validation sample. Pearson correlations were then computed between the actual and the estimated utilities.

Results

The mean correlations for each model in each cell of the design are reported in Table 2. As before, the correlations obtained for the constrained estimation (in parentheses) are superior to those obtained in the nonconstrained case (without parentheses). Hence, the discussion in the remainder of this section will also focus on the results obtained with the constrained procedures.

The part worth model performs best for medium and low noise levels, whatever the sample size, while the quadratic model is best only for high noise levels. Moreover, the improvement in predictive validity obtained with the part worth over the quadratic model reaches (0.989 − 0.968) 0.021, while the improvement obtained with the qua-

TABLE 2

Mean Pearson Correlations Obtained in the Four-Level Attributes Case[a]

			Noise Level		
			High	Medium	Low
Sample Size	High	L	0.863 (0.881)	0.901 (0.909)	0.908 (0.915)
		Q	0.871 (0.908)	0.964 (0.957)	0.980 (0.968)
		PW	0.833 (0.895)	0.962 (0.971)	0.986 (0.989)
	Medium	L	0.827 (0.855)	0.894 (0.902)	0.905 (0.913)
		Q	0.814 (0.881)	0.947 (0.948)	0.973 (0.964)
		PW	0.778 (0.862)	0.943 (0.959)	0.978 (0.982)
	Low	L	0.799 (0.826)	0.881 (0.892)	0.895 (0.907)
		Q	0.739 (0.855)	0.917 (0.934)	0.958 (0.957)
		PW	0.681 (0.831)	0.897 (0.935)	0.956 (0.972)

[a] L stands for linear, Q for quadratic and PW for part worth. Throughout the table, the entries in parentheses correspond to the results obtained with the constrained estimation procedures, while those without parentheses stand for the results obtained with OLS.

dratic model reaches (0.855 − 0.831) 0.024. The results obtained with the part worth model are comparatively better than those obtained in the three-level attribute case (Table 1). The main reason for these results is that the quadratic model does not perform well with part worth data. This is because the estimated quadratic functions which are constrained to be monotone (between the levels of interest) do not (and cannot) represent the part worth data well enough (more so with four levels than with three because of the greater nonlinearity).[7] Disaggregated results show that the loss in predictive validity reaches 0.074 when the quadratic model is used to estimate part worth data, and 0.077 when the part worth model is used to estimate linear data (while it goes up to 0.194 when the linear model is used to estimate part worth data).[8] Overall, it is best to use the part worth model with low and medium noise levels and the quadratic model for high noise levels (i.e., low predictive validity). Once again, one would not lose much if one used the quadratic or part worth model exclusively.

Conclusion

The simulation results show that constraining a function to be monotone (when known or expected) can improve substantially the predictive validity results. Hence, our preliminary set of recommendations is similar to the recommendations of Srinivasan, Jain and Malhotra (1983). First, constraints should be placed on the parameters of an attribute when it is obvious that the preference order of the attribute levels is monotone (e.g., miles per gallon). Second, respondents should be asked for the preference order of the levels of an attribute when monotonicity is not evident. This information can easily be included in a questionnaire and does not demand much additional effort on the part of the respondents. As argued earlier, the preference order will almost always be either monotone or of the ideal point type (see Figure 1). Whenever a preference order is monotone, one can then constrain the corresponding utility function. Note that one should also constrain the part worth utilities of a categorical attribute when the preference order of the levels is obvious, and ask respondents for their preference order when it is not obvious so as to constrain the part worth utilities accordingly. An estimation procedure which can constrain utility functions is LINMAP (Srinivasan and Shocker 1973).[9] One could also use an adaption of Cunia's procedure (1964).

Our main set of recommendations refers to the preference function to use in practice when it is known to be monotone. The most conservative (and least cumbersome) strategy involves using the quadratic or the part worth function for all attributes, but preferably the quadratic function. This is because one never loses much in terms of predictive validity with the quadratic function, and not much more with the part worth function, compared to the linear function. A less conservative strategy involves using (a) the quadratic function all the time with three-level attributes, (b) the part worth function with four-level attributes when the predictive validity is medium or high (i.e., when the adjusted R^2 is higher than about 0.7, or the fit at least about average with

[7] As a result, the predictive validity obtained on part worth data with the quadratic models tends to decrease when going from the nonconstrained to the constrained procedures. In fact, the average predictive validity obtained with the quadratic model decreases (Table 2) when the noise level is low (whatever the sample size) and when the noise level is medium and the sample size high.

[8] The disaggregated results are in Table 6 of (Cattin and Punj 1983). See Footnote 5.

[9] Note, however, that care should be taken in imposing constraints on any attribute which may be ecologically correlated with other attributes when these other attributes are left out. Price is such an attribute. "If the study included most of the relevant attributes of quality, and prestige is not an issue as far as the product/service is concerned, constrained estimation may be valuable. However, if relevant quality attributes have been left out and/or prestige is an important issue, the respondent may associate price with quality or prestige so that such a constrained estimation may actually decrease predictive validity" (Srinivasan et al. 1983).

nonmetric methods such as LINMAP), and (c) the quadratic function with four-level attributes when the predictive validity is low (i.e., when the adjusted R^2 is less than about 0.7).

Should one use quadratic or part worth functions with ideal point functions? One can use either one with three-level attributes (as argued earlier). With respect to four-level attributes, the results of Table 2 (obtained with nonconstrained estimation procedures) show that the part worth function outperforms the quadratic function in the low noise level cells, but by only 0.006 at the most. Hence, in the ideal point case, the part worth function would also outperform the quadratic function if the data are sufficiently different from the quadratic and the predictive validity is high. The most conservative strategy would involve using the quadratic function all the time. A less conservative strategy would involve using the part worth function only if the adjusted R^2 is high (i.e., about 0.9) or if the fit is good with nonmetric methods, and the quadratic function the rest of the time.[10]

[10]This paper was received December 1981 and has been with the authors for 2 revisions.

References

Cattin, Philippe (1982), "On the Estimation of Continuous Utility Functions in Conjoint Analysis," *Product Management and Quantitative Methods in Marketing*, J. D. Little and J. L. Chandon, eds., 249–267.

——— and Dick R. Wittink (1982), "Commercial Use of Conjoint Analysis: A Survey," *Journal of Marketing*, 46 (Summer), 44–53.

——— and Girish Punj (1983), "Evaluating Preference Models for Continuous Utility Functions in Conjoint Analysis," Working Paper No. 50-83, School of Business Administration, University of Connecticut.

Cunia, T. (1964), "Least Squares Estimates and Parabolic Regression with Restricted Location for the Stationary Point," *Journal of the American Statistical Association*, 59 (June), 564–591.

Green, Paul E. and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research*, 5 (September), 102–123.

Kuehn, Alfred A. and Ralph L. Day (1962), "Strategy of Product Quality," *Harvard Business Review*, 40, 100–110.

Pekelman, Dov and Subrata K. Sen (1979a), "Measurement and Estimation of Conjoint Utility Functions," *Journal of Consumer Research*, 5 (March), 263–271.

——— and ——— (1979b), "Improving Prediction in Conjoint Measurement," *Journal of Marketing Research*, 16 (May), 211–220.

Scott, Jerome E. and Peter Wright (1976), "Modeling an Organizational Buyer's Product Evaluation Strategy: Validity and Procedural Considerations," *Journal of Marketing Research*, 13 (August), 211–224.

Srinivasan, V. (1975), "Linear Programming Computational Procedures for Ordinal Regression," *Journal of the Association for Computing Machinery*, 23 (October), 475–487.

——— and Allan D. Shocker (1973), "Estimating the Weights for Multiple Attributes in a Composite Criterion Using Pairwise Judgments," *Psychometrika*, 38 (December), 473–493.

———, Arun K. Jain and Naresh K. Malhotra (1983), "Improving Predictive Power of Conjoint Analysis by Constrained Parameter Estimation," *Journal of Marketing Research*, 20 (November), forthcoming.