LOGISTIC REGRESSION RESPONSE FUNCTIONS WITH MAIN AND INTERACTION EFFECTS IN THE CONJOINT ANALYSIS

Amedeo DE LUCA[1,2]
PhD, University Professor,
Faculty of Economics, Department of Statistics Sciences,
The Catholic University, Milan, Italy

E-mail: amedeo.deluca@unicatt.it

Sara CIAPPARELLI[3]
BSc, Statistic and Economic Science,
Faculty of Economics, The Catholic University, Milan, Italy

E-mail:

Abstract: *In the Conjoint Analysis (COA) model proposed here - an extension of the traditional COA - the polytomous response variable (i.e. evaluation of the overall desirability of alternative product profiles) is described by a sequence of binary variables. To link the categories of overall evaluation to the factor levels, we adopt - at the aggregate level - a multivariate logistic regression model, based on a main and two-factor interaction effects experimental design.*
The model provides several overall desirability functions (aggregated part-worths sets), as many as the overall ordered categories are, unlike the traditional metric and non metric COA, which gives only one response function.
We provide an application of the model and an interpretation of the main and interactive effects.

Key words: *Conjoint analysis; Interaction effects; Multivariate Logistic Regression*

1. Introduction

Conjoint Analysis (COA) has become one of the most widely used quantitative tools in marketing research. COA has been developed and in use since the early 1970s and has aroused considerable interest as a major set of techniques for measuring consumers' trade-offs among multiattributed products and services [8][4].

Since that time many new developments in COA have been registered. COA comes in a variety of forms.

The purpose of this article is to give a new contribution to the problem of the conjoint measurement in order to quantify judgmental data (non metric COA) without resorting to scale adjustments to render the preference scale "metric".

The model proposed here is an extension of the traditional COA approach.

While in the traditional full-profile COA [8] the respondent expresses preferences by rating (ratings-based conjoint method) or ranking distinct product profiles, in our model we assume that the respondent expresses preferences by choosing the overall between *K* desirability categories for each of *S* hypothetical product profiles, chosen from a sample of respondents.

The proposed model also differs from the traditional methods relative to the case in which the response is on ordinal scale, as several authors - to study the relationship between the response and the explicative variables - apply the ordinal logistic regression or multinomial logit models for ordinal responses (adjacent-categories logits, continuation-ratio logit, cumulative logits or proportional odds model) or cumulative link models.

The proposed approach is different from the "Choice-Based Conjoint" Analysis (CBC) model in which the respondent expresses preferences by choosing concepts from sets of concept (discrete choice modelling).

A problem rarely considered in literature[1] [6] is the frequent renunciation of the measurement of the interaction effects between two or more product attributes [12].

Most applications of COA emphasize the main-effects models, as fewer data points are sufficient to fit such model at the individual level. But, such simplification reduces the predictive capacity of the model when its underlying utility or desirability functions must incorporate interaction effects between product attributes.

In the COA model proposed here to link the categories of overall evaluation to the factor levels, we adopt a multivariate logistic regression model [5] at the aggregate level [13]**,** based on a main and two-factor interaction effects experimental design.

The model provides (the novelty value in our approach) several overall desirability functions (aggregated part-worths sets), as many as the overall ordered categories are, unlike the traditional metric and non metric COA and CBC analysis, which give only one response function.

2. Estimation of Response Functions in the Conjoint Analysis via Multivariate Multiple Logistic Regression on Dummy Variables

In the COA approach proposed here it is assumed that the respondent evaluation of the *overall desirability*, that is to be expressed by a respondent sample on each of *S* hypothetical product profiles of the new product, consists only in one choice of the *K* desirability categories. To link the *overall* desirability (ordinal dependent variable *Y*, with modalities Y_k, $k = 1, 2, \ldots, K$) [4] with the levels of experimental factors (independent or predictor variables *X*) relative to the product, the summarizing vector of the choice probability of one of the *K* categories has been interpreted via a multivariate multiple logistic regression on dummy variables model.

The interpretative model of COA illustrated in this note is shown as complex enough from the statistical-mathematical point of view. For this reason, in the hope of making its reading easier, the methodology is presented in close connection with a concrete case, which can obviously be an example for further applications.

The reference application regards a survey carried out on a homogeneous sample of 100 insurance officers, who were asked a judgement of the *overall* desirability (*Y*) on $S = 24$ profiles of the insurance policy, described according to a factorial plan.

The evaluation scores have been classified in three categories: "undesirable", "desirable", "more desirable".

At this point we have just to link the overall desirability (Y_k, k = 1, 2, 3) to the factor levels (independent variables) related to the product (X_1 = "policy duration ", with levels: *5 years*, *8 years*; X_2 = "minimum denomination", levels: *2,500 euro*, *5,000 euro*; X_3 = "stock exchange index", levels: *Comit*, *Dow Jones*, *Nikkei*; X_4 = "service to expiry", levels: *paid-up capital*, *income for life*).

The reference frame is therefore to state the problem of studying an ordinal-scaled variable criterion, Y, "in function" to other predictor variables, X_1, X_2, X_3 , X_4, in the example given.

The model proposed concerns the full-profile COA and it is based on overall desirability categories chosen by a respondent sample, for each of S = 24 hypothetical product profiles. The total number of profiles or cards S, resulting from the total number of possible combinations of levels of the M = 4 attributes (X), equal respectively, to: 2, 2, 3, 2, constitute a *full-fact*orial experimental design of $2 \times 2 \times 3 \times 2 = 24$ stimuli. The focus of this study is to estimate the relationship between dependent and independent variables via a multivariate logistic regression model.

In the approach, for a given respondent j, we let y_{ksj} denote the desirability category k of the sth concept for the respondent j.

In terms of probabilities the effects of the factors express the variations of the probabilities p_{ks} - if k is the overall category - associated with the vector $\mathbf{z}_s$ corresponding to the combination s (s = 1, 2, ..., S) of levels of the M factor, as follows [10]:

$$p_{ks}(Y_k = 1 | \mathbf{z}_s) = \pi_k(\mathbf{z}_s) = \exp(\delta_k' \mathbf{z}_s) / [1 + \exp(\delta_k' \mathbf{z}_s)] \qquad (1)$$

where:

δ_k' is the unknown vector of regression coefficients of the predictor variables;

$\mathbf{z}_s$ is the vector of the dummy explanatory variables relative to the combination or profile s.

To estimate said probabilities $\pi_k(\mathbf{z}_s)$, we use an *aggregate* level model across the J homogeneous research respondents, whose evaluations, on each product profile, are considered J repeated observations.

To estimate the relationship between Y_k , k = 1, 2, ..., K (in our application K = 3) dependent variable (*overall* judgment category) and m = 1, 2, ..., M (in our application M = 4) qualitative independent variables (product attributes or factors X), with levels l = 1, 2, ..., l_m (in our application: l_1 = 2; l_2 = 2, l_3 = 3, l_4 = 2), the K overall categories (Y_k) are codified as K dummy variables (the scheme of such *disjunctive* binary coding [2] is displayed in Table 1a); also the independent variables are codified by Z binary variables; for each variable we have defined a set of 0-1 dummy variables $Z_l^{(m)}$ (l = 1, 2, ..., l_m) so that – for one m factor - $Z_l^{(m)}$ = 1 if category lth is observed, in all other cases $Z_l^{(m)}$ = 0 (Table 1b-1e).

Table 1a. Disjunctive binary coding of overall evaluations (Y_k) categories

Dummy variables / **Overall categories** (Y_k)	Y_1	Y_2	Y_3
k = 1 "undesirable"	1	0	0
k = 2 "desirable"	0	1	0
k = 3 "more desirable"	0	0	1

Table 1b-e. Disjunctive binary coding of factors: "policy duration ", "minimum denomination", "stock exchange index", "service to expiry"

b)

Predictor variables and levels (*l*)	**Dummy variables**	
Policy duration	$Z_1^{(1)}$	$Z_2^{(1)}$
5 years	1	0
8 years	0	1

c)

Predictor variables and levels (*l*)	**Dummy variables**	
Minimum denomination	$Z_1^{(2)}$	$Z_2^{(2)}$
2500 euros	1	0
5000 euros	0	1

d)

Predictor variables and levels (*l*)	**Dummy variables**		
Stock exchange index	$Z_1^{(3)}$	$Z_2^{(3)}$	$Z_3^{(3)}$
Comit	1	0	0
Dow Jones	0	1	0
Nikkei	0	0	1

e)

Predictor variables and levels(*l*)	**Dummy variables**	
Service to expiry	$Z_1^{(4)}$	$Z_2^{(4)}$
paid-up capital	1	0
income for life	0	1

The judgment evaluations are pooled across respondents (*pooled model*) and the novelty value in our approach is that one set of aggregated part-worths is estimated in connection with *each overall* category Y_k (see Table 3).

2.1. Identification of Univariate Multiple Logistic Regression Model

In the configured multivariate model, owing to the interrelationship between the *K* dependent variables (being the response categories on the *overall* exclusive and exhaustive within themselves), the *K*th equation can be drawn from the remaining $q = K-1$ equations. The $q = 2$ univariate logistic regression equations, *without* intercept, after transforming the dependent into a logit variable, are expressed as follows:

$$g_k(\mathbf{z}) = \ln[\pi_k(\mathbf{z})/(1-\pi_k(\mathbf{z}))]$$
$$= \mathbf{Z}\,\delta_k, \qquad k = 1, 2, \dots, q; \quad s = 1, 2, \dots, S \qquad (2)$$

where:

$g_k(\mathbf{z})$ is the logit transformation;

$\mathbf{Z}$ is $(SJ)\times(\sum_{m=1}^{M} l_m + \sum_{m=1}^{M}\sum_{p=1}^{P} l_{mp})$ fixed matrix of the experimental design (composed of *J* submatrices $\mathbf{Z}_j$ of indicator variables associated to the *S* combinations of the experimental plan; each $\mathbf{Z}_j$ has a dimension $S\times(\sum_{m=1}^{M} l_m + \sum_{m=1}^{M}\sum_{p=1}^{P} l_{mp})$ with elements $z_{ls}^{(m)}$ (dummy variable for the "*l*" level of the "*m*" factor in the combination "*s*"; $p = m + 1, m + 2, \dots, M$);

δ_k is a column vector of the unknown regression coefficients for the response function *k* with dimensions $(\sum_{m=1}^{M} l_m + \sum_{m=1}^{M}\sum_{p=1}^{P} l_{mp})$.

The $\mathbf{Z}$ matrix is expressed as follows (in the case here considered $M = 4$ with a number of categories, respectively, $l_1 = 2$; $l_2 = 2$; $l_3 = 3$; $l_4 = 2$):

$$
\mathbf{Z} = \left[\begin{array}{cccccccccccccccccc}
c_1 & | & z_{21}^{(1)} & z_{21}^{(2)} & z_{21}^{(3)} & z_{31}^{(3)} & z_{21}^{(4)} & | & z_{221}^{(12)} & z_{221}^{(13)} & z_{231}^{(13)} & z_{221}^{(14)} & z_{221}^{(23)} & z_{231}^{(23)} & z_{221}^{(24)} & z_{221}^{(34)} & z_{k321}^{(34)} \\
c_2 & | & z_{22}^{(1)} & z_{22}^{(2)} & z_{k22}^{(3)} & z_{32}^{(3)} & z_{22}^{(4)} & | & z_{222}^{(12)} & z_{222}^{(13)}v & z_{232}^{(13)} & z_{222}^{(14)} & z_{222}^{(23)} & z_{232}^{(23)} & z_{222}^{(24)} & z_{222}^{(34)} & z_{k322}^{(34)} \\
\vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
c_s & | & z_{2s}^{(1)} & z_{2s}^{(2)} & z_{2s}^{(3)} & z_{3s}^{(3)} & z_{2s}^{(4)} & | & z_{22s}^{(12)} & z_{22s}^{(13)} & z_{23s}^{(13)} & z_{22s}^{(14)} & z_{22s}^{(23)} & z_{23s}^{(23)} & z_{22s}^{(24)} & z_{22s}^{(34)} & z_{k32s}^{(34)} \\
\vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
c_S & | & z_{2S}^{(1)} & z_{2s}^{(2)} & z_{2S}^{(3)} & z_{3S}^{(3)} & z_{2S}^{(4)} & | & z_{22S}^{(12)} & z_{k22S}^{(13)} & z_{23S}^{(13)} & z_{22S}^{(14)} & z_{22S}^{(23)} & z_{23S}^{(23)} & z_{22S}^{(24)} & z_{22S}^{(34)} & z_{32S}^{(34)} \\
- & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - \\
c_1 & | & z_{21}^{(1)} & z_{21}^{(2)} & z_{21}^{(3)} & z_{31}^{(3)} & z_{21}^{(4)} & | & z_{221}^{(12)} & z_{221}^{(13)} & z_{231}^{(13)} & z_{221}^{(14)} & z_{221}^{(23)} & z_{231}^{(23)} & z_{221}^{(24)} & z_{221}^{(34)} & z_{321}^{(34)} \\
c_2 & | & z_{22}^{(1)} & z_{22}^{(2)} & z_{22}^{(3)} & z_{32}^{(3)} & z_{22}^{(4)} & | & z_{222}^{(12)} & z_{222}^{(13)} & z_{232}^{(13)} & z_{222}^{(14)} & z_{222}^{(23)} & z_{232}^{(23)} & z_{222}^{(24)} & z_{222}^{(34)} & z_{322}^{(34)} \\
\vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
c_s & | & z_{2s}^{(1)} & z_{2s}^{(2)} & z_{2s}^{(3)} & z_{3s}^{(3)} & z_{2s}^{(4)} & | & z_{22s}^{(12)} & z_{22s}^{(13)} & z_{23s}^{(13)} & z_{22s}^{(14)} & z_{22s}^{(23)} & z_{23s}^{(23)} & z_{22s}^{(24)} & z_{22s}^{(34)} & z_{32s}^{(34)} \\
\vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
c_S & | & z_{2S}^{(1)} & z_{2s}^{(2)} & z_{2S}^{(3)} & z_{3S}^{(3)} & z_{2S}^{(4)} & | & z_{22S}^{(12)} & z_{22S}^{(13)} & z_{23S}^{(13)} & z_{22S}^{(14)} & z_{22S}^{(23)} & z_{23S}^{(23)} & z_{22S}^{(24)} & z_{22S}^{(34)} & z_{32S}^{(34)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
c_1 & | & z_{21}^{(1)} & z_{21}^{(2)} & z_{21}^{(3)} & z_{31}^{(3)} & z_{21}^{(4)} & | & z_{221}^{(12)} & z_{221}^{(13)} & z_{231}^{(13)} & z_{221}^{(14)} & z_{221}^{(23)} & z_{231}^{(23)} & z_{221}^{(24)} & z_{221}^{(34)} & z_{321}^{(34)} \\
c_2 & | & z_{22}^{(1)} & z_{22}^{(2)} & z_{22}^{(3)} & z_{32}^{(3)} & z_{22}^{(4)} & | & z_{222}^{(12)} & z_{222}^{(13)} & z_{232}^{(13)} & z_{222}^{(14)} & z_{222}^{(23)} & z_{232}^{(23)} & z_{222}^{(24)} & z_{222}^{(34)} & z_{322}^{(34)} \\
\vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
c_s & | & z_{2s}^{(1)} & z_{2s}^{(2)} & z_{2s}^{(3)} & z_{3s}^{(3)} & z_{2s}^{(4)} & | & z_{22s}^{(12)} & z_{22s}^{(13)} & z_{23s}^{(13)} & z_{22s}^{(14)} & z_{22s}^{(23)} & z_{23s}^{(23)} & z_{22s}^{(24)} & z_{22s}^{(34)} & z_{32s}^{(34)} \\
\vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
c_S & | & z_{2S}^{(1)} & z_{2s}^{(2)} & z_{2S}^{(3)} & z_{3S}^{(3)} & z_{2S}^{(4)} & | & z_{22S}^{(12)} & z_{22S}^{(13)} & z_{23S}^{(13)} & z_{22S}^{(14)} & z_{22S}^{(23)} & z_{23S}^{(23)} & z_{22S}^{(24)} & z_{22S}^{(34)} & z_{32S}^{(34)}
\end{array}\right]
\begin{array}{l}
\blacktriangleright \quad j = 1 \\ \\ \\ \\ \\ \\ \\
\blacktriangleright \quad j = 2 \\ \\ \vdots \\ \\ \\ \\ \\
\blacktriangleright \quad j = 100 \\ \\ \\ \\ \\ \\
\end{array}
$$

To resolve the linear dependency between the independent variables the model is reparametrized [14] using $z_1^{(m)}$ as a reference category and the algebraic form response functions, *with intercept* (following the omission of one column from each attribute, included as a reference level (*baseline*), correspondent to the first category ($z_{1s}^{(m)}$) with main and first-order interaction effects) is:

$$
g_k(\tilde{\mathbf{z}}_s) = \tilde{\delta}_{k0} + \sum_{m=1}^{M} \sum_{l=2}^{l_m} \tilde{\delta}_{kl}^{(m)} \tilde{z}_{ls}^{(m)} + \sum_{m=1}^{M} \sum_{l=2}^{l_m} \sum_{h=2}^{h_p} \tilde{\delta}_{klh}^{(m,p)} \tilde{z}_{l\,h\,s}^{(m,p)} + e_{ks} \qquad (3)
$$

$k = 1, \ldots, q;$ $s = 1, 2, \ldots, S;$ $h = 1, 2, \ldots, h_m;$ $p = m + 1, m + 2, \ldots, M;$

where:

$\tilde{g}_k(\tilde{\mathbf{z}}_s)$ is the logit of the *s*th profile with regard to the *k*th dependent variable;

$\tilde{\delta}_{k0}$ is the constant term;

$\tilde{\delta}_{kl}^{(m)}$ is the unknown regression coefficient for the *l*th level of the *m* factor;

$\tilde{z}_{ls}^{(m)}$ is the dummy variable for the *l*th level of the *m* factor in the combination *s*;

$\tilde{z}_{lhs}^{(m,p)}$ is the dummy variable for the interaction between the *l*th level of the *m* factor and the *h*th level of the *p* factor in the combination *s*;

e_{ks} is the error term pertinent to the *s*th stimulus.

2.2. Identification of Univariate Multiple Logistic Regression on Dummy variables

The *q* equations (3), *with intercept* $\tilde{\delta}_{k0}$ correspondent to the first category ($\tilde{z}_{1s}^{(m)}$) of each factor *m*, are:

$$g_k(\tilde{\mathbf{z}}) = \ln[\pi_k(\tilde{\mathbf{z}})/(1-\pi_k(\tilde{\mathbf{z}})] = \tilde{\mathbf{Z}}\,\tilde{\delta}_k, \qquad k = 1, 2, \ldots, q, \tag{4}$$

where:

$g_k(\tilde{\mathbf{z}})$ is the *logit transformation*;

$\tilde{\mathbf{Z}}$ is a fixed matrix (the design matrix below equation (3)) and has dimensions $(SJ) \times (1 + \sum_{m=2}^{M} l_m + \sum_{m=2}^{M}\sum_{p=2}^{P} l_{mp})$;

$\tilde{\delta}_k$ is a column vector of the regression coefficients.

The *q* equations $g_k(\tilde{\mathbf{z}}_s)$ can be expressed compactly as follows [5]:

$$\mathbf{g}^* = \tilde{\mathbf{Z}}^* \tilde{\delta}^*, \tag{5}$$

where:

$\mathbf{g}^*$ is a compound vector (vec) of *q* column vectors $g_k(\tilde{\mathbf{z}})$;

$\tilde{\mathbf{Z}}^*$ is a $[q(JS)] \times [q(1 + \sum_{m=1}^{M} l_m + \sum_{m=2}^{M}\sum_{p=2}^{P} l_{mp})]$ *square* compound diagonal matrix, containing $q \times q$ submatrices (each one of dimensions $(SJ \times [(1 + \sum_{m=1}^{M} l_m + \sum_{m=2}^{M}\sum_{p=2}^{P} l_{mp})]$ of which the *q* $\tilde{\mathbf{Z}}$ submatrices disposed along the principal diagonal (all equal) give the independent indicator variables relative to the various equations in the column, while the remaining submatrices are compounds of zero elements;

$\tilde{\delta}^*$ is a compound vector of the *q* column vectors $\tilde{\delta}_k$ of the regression coefficients, each with dimensions $(1 + \sum_{m=1}^{M} l_m + \sum_{m=2}^{M}\sum_{p=2}^{P} l_{mp}) \times 1$.

To estimate the (5) multivariate model parameters we need to consider the following variance-covariance matrix Φ, between the Y_k, with elements Var(Y_{ksj}) = $p_{ksj}(1 - p_{ksj})$ [11], where p_{ksj} is the probability for a *j* respondent to choose the *k* category for the *s* combination, and Cov(Y_{ksj}, Y_{qsj}) = - $p_{ksj} \cdot p_{qsj}$, defined as follows [3]:

$$\Phi = \left[\begin{array}{ccc|ccc|ccc}
p_{111}(1-p_{111}) & \cdots & 0 & -p_{111}p_{211} & \cdots & 0 & \ldots -p_{111}p_{q11} & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & p_{1SJ}(1-p_{1SJ}) & 0 & \cdots & -p_{1SJ}p_{2SJ} & 0 & \cdots & -p_{1SJ}p_{2SJ} \\
--- & --- & --- & --- & --- & --- & --- & --- & --- \\
-p_{211}p_{111} & \cdots & 0 & p_{211}(1-p_{211}) & \cdots & 0 & \ldots -p_{211}p_{q11} & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & -p_{2SJ}p_{1SJ} & 0 & \cdots & p_{2SJ}(1-p_{2SJ}) & 0 & \cdots & -p_{2SJ}p_{qSJ} \\
--- & --- & --- & --- & --- & --- & --- & --- & --- \\
-p_{q11}p_{111} & \cdots & 0 & -p_{q11}p_{211} & \cdots & 0 & \ldots p_{q11}(1-p_{q11}) & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & -p_{qSJ}p_{1SJ} & 0 & \cdots & -p_{qSJ}p_{2SJ} & 0 & \cdots & p_{qSJ}(1-p_{qSJ})
\end{array}\right]$$

The estimates of the Φ matrix elements are calculated on the basis of estimations $\hat{p}_{ksj}$ (k = 1, 2, ..., q) [5], obtained by performing logistic regression analysis separately on each dependent variable, using the maximum likelihood method to each equation (3).
To estimate the multivariate logistic regression model (5) we minimize the following mathematical expression:

$$F = (\mathbf{g}^* - \tilde{\mathbf{Z}}^* \tilde{\delta}^*)' \hat{\Phi}^{-1} (\mathbf{g}^* - \tilde{\mathbf{Z}}^* \tilde{\delta}^*); \qquad (6)$$

where $\hat{\Phi}^{-1}$ is the inverse matrix of the $\hat{\Phi}$, $\tilde{\mathbf{g}}^*$ is a compound vector (vec) of q column vectors $\tilde{g}_k(\tilde{\mathbf{z}}_s)$.

3. The Application of the Proposed Model

The model was applied to the overall desirability evaluations expressed on the K = 3 categories: "undesirable", "desirable", "more desirable", by a sample of J = 100 insurance officers (homogeneous respondents) on S = 24 profiles of the insurance policy.

The M = 4 experimental factors (attributes) and levels were: X_1 = "policy duration" (with levels: 5, 8 years); X_2 = "minimum denomination" (2,500 €, 5,000 €); X_3 = "stock exchange index" (Comit, Dow Jones, Nikkei); X_4 = "service to expiry" (paid-up capital, income for life).

Table 2 quotes the complete factorial design with restricted casualisation submitted to the interviewed. To estimate the parameters (part-worths) of the response functions of the (6), the Constrained Non Linear Regression (CNLR) [7] program of the SPSS software was used [15].

These estimates of the regression coefficient values, which are equal to the constant term plus the corresponding parameters, are given in Table 3 (Appendix 1); $\tilde{\delta}_{kl}^{(m)}$ is the coefficient relative to the k equation, m factor and l level. In Table 3, the positive signs of the coefficients indicate that the respective response variables increase in relation to the level in the single product factor and vice-versa.

Table 2. Full factorial design with restricted casualisation of four factors and factor levels of the index-linked life policy

Stimulus (*s*)	Policy duration (years) (X_1)	Minimum denomination (euro) (X_2)	Stock exchange index (X_3)	Service to expiry (X_4)
1	5	2500	Nikkei	paid-up capital
16	5	5000	Comit	paid-up capital
9	8	5000	Dow Jones	paid-up capital
18	8	5000	Comit	paid-up capital
12	5	5000	Dow Jones	paid-up capital
19	8	2500	Nikkei	paid-up capital
10	8	2500	Comit	paid-up capital
24	8	2500	Dow Jones	paid-up capital
2	5	2500	Comit	paid-up capital
15	5	5000	Nikkei	paid-up capital
5	8	5000	Nikkei	paid-up capital
17	5	2500	Dow Jones	paid-up capital
6	8	5000	Comit	income for life
21	8	2500	Dow Jones	income for life
7	5	5000	Comit	income for life
23	5	2500	Nikkei	income for life
8	5	5000	Dow Jones	income for life
14	5	2500	Comit	income for life
11	5	2500	Dow Jones	income for life
20	8	5000	Nikkei	income for life
4	8	2500	Nikkei	income for life
13	8	5000	Dow Jones	income for life
3	8	2500	Comit	income for life
22	5	5000	Nikkei	income for life

4. Meaning of the Proposed Interpretative Model and Some Empirical Results

Out of a reading of the coefficients (*effects*) in Table 3, we can see the modalities of the factors that contribute to the increase/decrease of the $\hat{p}_k$ (k = 1, 2, 3) values, and consequently the relative importance of each attribute as well as which levels of each attribute are most preferred.

Table 3 points out that the main effect values related to the first equation, associated to the "undesirable" global evaluation category, are of opposite algebraic sign – if the second modality of the second factor and the second modality of the fourth are excepted – to those correspondent to the second and third equation (associated, respectively, to the "desirable" and "more desirable" judgment categories).

This allows verifying the basic coherence of the results of the model (at least with regard to the main effects).

With regard to the interactive effects, less univocal relations on the algebraic signs of the part-worth coefficients are noticed and we find out what follows.

In the first equation of the model (k = 1) we observe five negative coefficients, against three in the second (k = 2) and the third (k = 3) equation; moreover between the positive coefficients of the first equation we notice higher values than those of the other two equations (1,052 in connection to the third modality of the factor 3 and the second category of factor 4; 0,935 in connection to the second category of factor 3 and 4).

In the third equation we see the lowest negative coefficients (-1,153 in connection with the third modality of factor 3 and the second modality of factor 4; -0,640 in connection with the second category of factors 1 and 4).

In order to empirically assess the predictive capacity of the estimated model, see Table 4 (Appendix 2) shows the probabilities estimated $\hat{p}_k$, for $k = 1, 2, 3$, by (6) for all the modality combinations (experimental conditions) of the explanatory variables and the corresponding values of the observed proportions. We observe a satisfactory model fitting, as the predicted probabilities turn out to be very near the corresponding proportions for all the modality combinations of the experimental design.

5. Final Remarks

The proposed model, at the aggregate level offers the prospect of more accurate estimation, unlike the traditional conjoint methods which estimate part-worth utilities at the individual level, moreover we can argue that aggregate analysis could permit the estimation of subtle interaction effects due to its ability to leverage a great deal of data across respondents.

Besides these positive features the model here proposed provides the following remarkable advantages: 1) it allows to estimate the main and two-factor interaction effects; 2) the inclusion of the interactive effects increased the predictability of the model itself; 3) the use of the probability $\hat{p}_{ks}$ as an average response, which does not require scale adjustments (by means of the traditional multidimensional scaling methods) to render the preference scale "metric"; 4) the estimate of one set of aggregated part-worths in connection with each overall category *k* and 5) a cross-check of the effects of the attribute levels on the different k categories. This allows us to verify the basic coherence of the results of the model, unlike the classical approaches (metric and non metric COA) adopted in literature.

6. Discussion

In the COA there are two critical questions: how many attributes and which ones are to be included in a study of conjoint analysis? On of the most fundamental problems in COA is reducing the number of profiles that need to be evaluated by respondents.

In our application we used the full factorial design; the covariates are four and their levels are from two to three. But this design is workable when the attributes and levels, considered relevant, are numerically very limited.

When the number of the covariates and/or the number of the levels are very large, it is necessary to have recourse to same types of fractional factorial designs in which not all the experimental combinations are constructed.

The most common type of fractional factorial regards the case in which all main effect and all two-factor interactions can be estimated; we assume that higher interactions (involving three or more factors) are negligible and can be ignored.

In case a relatively large number of factors (e.g., nine or ten) are involved, the traditional type of fractional factorial (for example, "Resolution V design"), is used.

But, when we have twelve or more factors, each involving two to six levels, even the more traditional fractional factorial designs would require too many experimental combinations to estimate all main and two-factor interaction effects. In this case we can make use of other newer classes of fractional factorial designs (for example, Compromise design; Interactive Conjoint Scaling - ICONS [9]).

Then we should note that in case of uncertainty about the covariates to be included and the number of their levels, we can resort, for example, to CBC analysis. Respondents are shown a set of products in full-profiles and asked to indicate which one they would like to purchase. CBC can measure up to six attributes with nine levels each, with the inclusion of attribute interactions, at the aggregate level, or group level.

When the number of attributes is more than six we can use to Adaptive Conjoint Analysis (ACA). ACA is a computer-administered interactive method designed for situation in which the number of attributes exceeds what can be reasonably done with more traditional methods (such as CBC). ACA focuses on the attributes that are most relevant to the respondent and avoids overload by focusing on just a few attributes at a time.

ACA System can handle a number of attributes as large as 30, up to 15 levels per attribute, but it is a main-effects model that not include the attribute interactions.

In conclusion, it makes little sense to argue over which form of COA model is the overall best approach, as we should consider the different research situations for the choice of the specific model.

References

1. Carmone, F. J. and Green, P. E. **Model misspecification in multiattribute parameter estimation,** Journal of Marketing Research, 18, 1981, pp. 87-93
2. De Luca, A. **Un modello di misurazione della customer satisfaction con codifica binaria additiva dei predittori ordinali,** in Zanella, A. (ed.) "Valutazione della qualità e customer satisfaction: il ruolo della Statistica", Milano: Vita e Pensiero, 2000, pp. 275-290
3. De Luca, A., Zanella, A. and Cantaluppi, G. **Un modello di valutazione della customer satisfaction con variabile risposta politomica,** Statistica Applicata, 4, 2004, pp. 563-598
4. De Luca, A. **A logit model with a variable response and predictors on an ordinal scale to measure customer satisfaction,** Quality and Reliability Engineering International, 22, 2006, pp. 591-602
5. De Luca, A. and Ciapparelli, S. **Multivariate Logistic Regression for the Estimate of Response Functions in the Conjoint Analysis,** in "MTISD 2008: Methods, Models and Information Technologies for Decision Support Systems", Lecce, 18 – 20 September, 2008, pp. 55-58
6. Fenwick, I. **A problem in quantifying consumer trade-offs,** Journal of the Market research Society, 21, 1979, pp. 206-210
7. Gill, P.E., Murray, W. and Wrigth, M.H. **Pratical optimization,** San Francisco: Academic Press, 1981
8. Green, P.E. and Rao, V.R. **Conjoint measurement for quantifying judgmental data,** Journal of Marketing Research, 8, 1971, pp. 355-363
9. Green, P.E., Carroll, J.D. and Carmone, F.J. **Some new types of fractional factorial designs for marketing experiments,** Research in Marketing, 1, 1978, pp. 99-122
10. Hosmer, D.W. and Lemeshow, S. **Applied Logistic Regression,** New York: John Wiley, 1989
11. Kotz, S. and Johnson, N. L. **Encyclopedia of statistical sciences, Volume 5,** New York: John. Wiley, 1985
12. Molteni, L. and Manoforte, R. **La conjoint analysis e il problema delle interazioni fra gli attributi: un'evidenza empirica,** Milano: Liuc Papers, 1998
13. Moore, W.L. **Levels of aggregation in conjoint analysis: an empirical comparison,** Journal of Marketing Research, 17, 1980, pp. 516-523

14. Suits, D.B. **Use of dummy variables in regression equations,** Journal of the American Statistical Association, 52, 1959, pp. 548-551
15. **Constrained Non Linear Regression- CNLR,** Chicago: SPSS Inc., 2002

Appendix 1

Table 3. Estimates of three set of the aggregated part-worths utilities of the COA model

Overall category	Estimated coefficient of the first equation		Overall category	Estimated coefficient of the second equation		Overall category	Estimated coefficient of the third equation	
	$\tilde{c}_1$	-0,093		$\tilde{c}_2$	0,409		$\tilde{c}3$	-1,955
	$\tilde{\delta}_{12}^{(1)}$	0,126		$\tilde{\delta}_{22}^{(1)}$	-0,073		$\tilde{\delta}_{32}^{(1)}$	-0,136
	$\tilde{\delta}_{12}^{(2)}$	0,058		$\tilde{\delta}_{22}^{(2)}$	0,181		$\tilde{\delta}_{32}^{(2)}$	-0,704
	$\tilde{\delta}_{12}^{(3)}$	-2,165		$\tilde{\delta}_{22}^{(3)}$	0,528		$\tilde{\delta}_{32}^{(3)}$	1,446
	$\tilde{\delta}_{13}^{(3)}$	2,267		$\tilde{\delta}_{23}^{(3)}$	0,551		$\tilde{\delta}_{33}^{(3)}$	1,458
	$\tilde{\delta}_{12}^{(4)}$	0,403		$\tilde{\delta}_{22}^{(4)}$	-0,465		$\tilde{\delta}_{32}^{(4)}$	0,041
	$\tilde{\delta}_{122}^{(12)}$	-0,070		$\tilde{\delta}_{222}^{(12)}$	-0,108		$\tilde{\delta}_{322}^{(12)}$	0,547
"undesirable"	$\tilde{\delta}_{122}^{(13)}$	-0,120	"desirable"	$\tilde{\delta}_{222}^{(13)}$	0,009	"more desirable"	$\tilde{\delta}_{322}^{(13)}$	0,201
	$\tilde{\delta}_{123}^{(13)}$	-0,026		$\tilde{\delta}_{223}^{(13)}$	-0,122		$\tilde{\delta}_{323}^{(13)}$	0,305
	$\tilde{\delta}_{122}^{(14)}$	0,113		$\tilde{\delta}_{222}^{(14)}$	0,111		$\tilde{\delta}_{322}^{(14)}$	-0,640
	$\tilde{\delta}_{122}^{(23)}$	-0,019		$\tilde{\delta}_{222}^{(23)}$	0,207		$\tilde{\delta}_{322}^{(23)}$	0,256
	$\tilde{\delta}_{123}^{(23)}$	-0,131		$\tilde{\delta}_{223}^{(23)}$	0,394		$\tilde{\delta}_{323}^{(23)}$	0,087
	$\tilde{\delta}_{122}^{(24)}$	0,152		$\tilde{\delta}_{222}^{(24)}$	-0,477		$\tilde{\delta}_{322}^{(24)}$	0,767
	$\tilde{\delta}_{122}^{(34)}$	0,935		$\tilde{\delta}_{222}^{(34)}$	0,747		$\tilde{\delta}_{322}^{(34)}$	-0,156
	$\tilde{\delta}_{132}^{(34)}$	1,052		$\tilde{\delta}_{232}^{(34)}$	0,503		$\tilde{\delta}_{332}^{(34)}$	-1,153

Appendix 2

Table 4. Comparison of the predicted probabilities, estimated by the COA model, and the corresponding proportions for all the modality combinations of the experimental design

Experimental conditions	Proportions observed	Predicted pro-babilities $\pi_1(\mathbf{z}_s)$	Proportions observed	Predicted pro-babilities $\pi_2(\mathbf{z}_s)$	Proportions observed	Predic-ted pro-babilities $\pi_3(\mathbf{z}_s)$
$z_1^{(1)}=1; z_1^{(2)}=1; z_1^{(3)}=1; z_1^{(4)}=1$	0,45	0,48	0,45	0,40	0,10	0,12
$z_1^{(1)}=1; z_2^{(2)}=1; z_2^{(3)}=1; z_1^{(4)}=1; z_{22}^{(2,3)}=1$	0,10	0,10	0,61	0,62	0,29	0,28
$z_2^{(1)}=1; z_2^{(2)}=1; z_3^{(3)}=1; z_1^{(4)}=1; z_{22}^{(1,2)}=1; z_{23}^{(1,3)}=1; z_{23}^{(2,3)}=1$	0,08	0,08	0,60	0,60	0,32	0,40
$z_2^{(1)}=1; z_2^{(2)}=1; z_2^{(3)}=1; z_1^{(4)}=1; z_{22}^{(1,2)}=1; z_{22}^{(1,3)}=1; z_{22}^{(2,3)}=1$	0,10	0,09	0,60	0,58	0,30	0,41
$z_1^{(1)}=1; z_2^{(2)}=1; z_3^{(3)}=1; z_1^{(4)}=1; z_{23}^{(2,3)}=1$	0,05	0,08	0,67	0,67	0,28	0,25
$z_2^{(1)}=1; z_1^{(2)}=1; z_1^{(3)}=1; z_1^{(4)}=1$	0,52	0,51	0,34	0,38	0,14	0,11
$z_2^{(1)}=1; z_1^{(2)}=1; z_2^{(3)}=1; z_1^{(4)}=1; z_{22}^{(1,3)}=1$	0,08	0,10	0,54	0,51	0,38	0,39
$z_2^{(1)}=1; z_1^{(2)}=1; z_3^{(3)}=1; z_1^{(4)}=1; z_{23}^{(1,3)}=1$	0,12	0,09	0,48	0,49	0,40	0,42
$z_1^{(1)}=1; z_1^{(2)}=1; z_2^{(3)}=1; z_1^{(4)}=1$	0,10	0,09	0,50	0,53	0,40	0,38
$z_1^{(1)}=1; z_2^{(2)}=1; z_1^{(3)}=1; z_1^{(4)}=1$	0,50	0,49	0,43	0,44	0,07	0,07
$z_2^{(1)}=1; z_2^{(2)}=1; z_1^{(3)}=1; z_1^{(4)}=1; z_{22}^{(1,2)}=1$	0,51	0,51	0,41	0,40	0,08	0,10
$z_1^{(1)}=1; z_1^{(2)}=1; z_3^{(3)}=1; z_1^{(4)}=1$	0,10	0,09	0,55	0,54	0,35	0,38
$z_2^{(1)}=1; z_2^{(2)}=1; z_2^{(3)}=1; z_2^{(4)}=1; z_{22}^{(1,2)}=1; z_{22}^{(1,3)}=1; z_{22}^{(1,4)}=1; z_{22}^{(2,3)}=1; z_{22}^{(2,4)}=1; z_{22}^{(3,4)}=1$	0,35	0,34	0,54	0,56	0,11	0,15
$z_2^{(1)}=1; z_1^{(2)}=1; z_3^{(3)}=1; z_2^{(4)}=1; z_{23}^{(1,3)}=1; z_{24}^{(1,4)}=1; z_{32}^{(3,4)}=1$	0,31	0,33	0,54	0,52	0,15	0,11
$z_1^{(1)}=1; z_2^{(2)}=1; z_3^{(3)}=1; z_2^{(4)}=1; z_{22}^{(2,3)}=1$	0,31	0,33	0,60	0,58	0,09	0,15
$z_1^{(1)}=1; z_1^{(2)}=1; z_1^{(3)}=1; z_2^{(4)}=1$	0,59	0,58	0,26	0,29	0,15	0,13
$z_1^{(1)}=1; z_2^{(2)}=1; z_3^{(3)}=1; z_2^{(4)}=1; z_{23}^{(2,3)}=1; z_{22}^{(2,4)}=1; z_{32}^{(3,4)}=1$	0,32	0,30	0,57	0,57	0,11	0,19
$z_1^{(1)}=1; z_1^{(2)}=1; z_2^{(3)}=1; z_2^{(4)}=1; z_{22}^{(3,4)}=1$	0,27	0,29	0,62	0,60	0,11	0,12
$z_1^{(1)}=1; z_1^{(2)}=1; z_3^{(3)}=1; z_2^{(4)}=1; z_{32}^{(3,4)}=1$	0,29	0,29	0,53	0,55	0,18	0,17
$z_2^{(1)}=1; z_2^{(2)}=1; z_1^{(3)}=1; z_2^{(4)}=1; z_{22}^{(1,2)}=1; z_{22}^{(1,4)}=1; z_{22}^{(2,4)}=1$	0,65	0,67	0,22	0,22	0,13	0,11
$z_2^{(1)}=1; z_1^{(2)}=1; z_1^{(3)}=1; z_2^{(4)}=1; z_{22}^{(1,4)}=1$	0,63	0,63	0,33	0,30	0,04	0,06
$z_2^{(1)}=1; z_2^{(2)}=1; z_3^{(3)}=1; z_2^{(4)}=1; z_{22}^{(1,2)}=1; z_{23}^{(1,3)}=1; z_{22}^{(1,4)}=1$	0,34	0,34	0,52	0,52	0,14	0,20
$z_2^{(1)}=1; z_1^{(2)}=1; z_1^{(3)}=1; z_2^{(4)}=1; z_{22}^{(1,3)}=1; z_{22}^{(1,4)}=1; z_{22}^{(3,4)}=1$	0,33	0,31	0,59	0,61	0,08	0,07
$z_1^{(1)}=1; z_2^{(2)}=1; z_1^{(3)}=1; z_2^{(4)}=1; z_{22}^{(2,4)}=1$	0,63	0,63	0,24	0,24	0,13	0,14

Appendix 3

Indicator variables corresponding to the overall variable Y of an interviewed generic and indicator variables relative to the product attributes (X_1, X_2, X_3, X_4), whit reference to the full factorial design with restricted casualisation of the S = 24 experimental stimuli (s = 1, 2, …, 24)

s:	Y_1	Y_2	Y_3	Z_{11}	Z_{12}	Z_{21}	Z_{22}	Z_{31}	Z_{32}	Z_{33}	Z_{41}	Z_{42}
1	0	1	0	1	0	1	0	0	0	1	1	0
16	1	0	0	1	0	0	1	1	0	0	1	0
9	0	1	0	0	1	0	1	0	1	0	1	0
18	1	0	0	0	1	0	1	1	0	0	1	0
12	0	0	1	1	0	0	1	0	1	0	1	0
19	0	1	0	0	1	1	0	0	0	1	1	0
10	0	1	0	0	1	1	0	0	0	1	1	0
24	0	1	0	0	1	1	0	0	1	0	1	0
2	1	0	0	1	0	1	0	1	0	0	1	0
15	0	1	0	1	0	0	1	0	0	1	1	0
5	0	1	0	0	1	0	1	0	0	1	1	0
17	0	0	1	1	0	1	0	0	1	0	1	0
6	1	0	0	0	1	0	1	1	0	0	0	1
21	1	0	0	0	1	1	0	0	1	0	0	1
7	1	0	0	1	0	0	1	1	0	0	0	1
23	0	0	1	1	0	1	0	0	0	1	0	1
8	0	1	0	1	0	0	1	0	1	0	0	1
14	1	0	0	1	0	1	0	1	0	0	0	1
11	1	0	0	1	0	1	0	0	1	0	0	1
20	0	1	0	0	1	0	1	0	0	1	0	1
4	1	0	0	0	1	1	0	0	0	1	0	1
13	1	0	0	0	1	0	1	0	1	0	0	1
3	1	0	0	0	1	1	0	1	0	0	0	1
22	0	0	1	1	0	0	1	0	0	1	0	1

[1] **He teaches** Market analysis 1, 2, and Planning and Market Analysis.
Over 170 Published papers, 12 books.
Research interests fields: Data analysis; Design and Analysis of Experiments; Marketing models; Statistics and econometrics; Conjoint analysis models; Customer satisfaction models; Data Mining.
In the past has taught in many university courses (including the Master of Managerial Sciences, Transylvania University, Brasov, Romania).
He is a member of the Permanent Working Group "Statistics for the Evaluation and Quality of Services (SVQS)", Italian Statistical Society.
He is a member of the Scientific Committee of EJASA.

[2] A. De Luca developed the model and its probability interpretation, and also wrote this note; S. Ciapparelli was responsable for the computer processing of the application.

[3] **Sara Ciapparelli**
Degree in Statistic and Economic Science at Faculty of Economy, Catholic University Milan (graduated in 2003).
After 5 years in an International Market Research Company in charge of sample design, since 2007, Risk Analyst in a Bank.
Five published conference technical papers about ordinal logistic regression for the Estimate of Response Functions in the Conjoint Analysis.
Research interest fields:
- Customer Satisfaction

- Economic and Statistic Models
- Credit Score Models

[4] Codification of references in text

[1]	Carmone, F. J. and Green, P. E. **Model misspecification in multiattribute parameter estimation,** Journal of Marketing Research, 18, 1981, pp. 87-93
[2]	De Luca, A. **Un modello di misurazione della customer satisfaction con codifica binaria additiva dei predittori ordinali,** in Zanella, A. (ed.) "Valutazione della qualità e customer satisfaction: il ruolo della Statistica", Milano: Vita e Pensiero, 2000, pp. 275-290
[3]	De Luca, A., Zanella, A. and Cantaluppi, G. **Un modello di valutazione della customer satisfaction con variabile risposta politomica,** Statistica Applicata, 4, 2004, pp. 563-598
[4]	De Luca, A. **A logit model with a variable response and predictors on an ordinal scale to measure customer satisfaction,** Quality and Reliability Engineering International, 22, 2006, pp. 591-602
[5]	De Luca, A. and Ciapparelli, S. **Multivariate Logistic Regression for the Estimate of Response Functions in the Conjoint Analysis,** in "MTISD 2008: Methods, Models and Information Technologies for Decision Support Systems", Lecce, 18 – 20 September, 2008, pp. 55-58
[6]	Fenwick, I. **A problem in quantifying consumer trade-offs,** Journal of the Market research Society, 21, 1979, pp. 206-210
[7]	Gill, P.E., Murray, W. and Wrigth, M.H. **Pratical optimization,** San Francisco: Academic Press, 1981
[8]	Green, P.E. and Rao, V.R. **Conjoint measurement for quantifying judgmental data,** Journal of Marketing Research, 8, 1971, pp. 355-363
[9]	Green, P.E., Carroll, J.D. and Carmone, F.J. **Some new types of fractional factorial designs for marketing experiments,** Research in Marketing, 1, 1978, pp. 99-122
[10]	Hosmer, D.W. and Lemeshow, S. **Applied Logistic Regression,** New York: John Wiley, 1989
[11]	Kotz, S. and Johnson, N. L. **Encyclopedia of statistical sciences, Volume 5,** New York: John. Wiley, 1985
[12]	Molteni, L. and Manoforte, R. **La conjoint analysis e il problema delle interazioni fra gli attributi: un'evidenza empirica,** Milano: Liuc Papers, 1998
[13]	Moore, W.L. **Levels of aggregation in conjoint analysis: an empirical comparison,** Journal of Marketing Research, 17, 1980, pp. 516-523
[14]	Suits, D.B. **Use of dummy variables in regression equations,** Journal of the American Statistical Association, 52, 1959, pp. 548-551
[15]	**Constrained Non Linear Regression- CNLR,** Chicago: SPSS Inc., 2002