Center
for
Economic Research

No. 2000-119

COMBINING AND COMPARING CONSUMERS' STATED PREFERENCE RATINGS AND CHOICE RESPONSES

By Bernard Conlon, Benedict G.C. Dellaert and
Arthur van Soest

December 2000

ISSN 0924-7815

Combining and Comparing Consumers' Stated Preference Ratings and Choice Responses

Bernard Conlon[1], Benedict G.C. Dellaert[2] and Arthur van Soest[3]

[1]CentER, [2]Department of Marketing and [3]Department of Econometrics
Faculty of Economics and Business Administration
Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

November 2000

JEL Codes: C35, M31

Key words: stated preference data, random coefficients, simulated ML

Abstract

In this study we develop and test an econometric model for combining choice and preference ratings data collected from the same set of individuals. Choice data are modeled using a multinomial logit framework, while preference data are modeled using an ordered response equation. Individual heterogeneity is allowed for via random coefficients providing a link between the choice and ratings data. Parameters are estimated by Simulated Maximum Likelihood. An application of the model to consumer yoghurt choice in The Netherlands found that ratings based preference estimates differ significantly from choice based estimates, but the correlation between random coefficients driving the two is very strong.

Acknowledgments
The authors thank Robert Bartels, Joffre Swait and Bas Donkers for useful comments and suggestions. The second author's research on this project was funded by the Dutch Science Foundation (NWO-ESR). A Tilburg University, Faculty of Economics and Business Administration grant for data collection through the CentER*data*-panel is gratefully acknowledged.

Combining and Comparing Consumers' Stated Preference Ratings and Choice Responses

Key words: stated preference data, random coefficients, simulated ML

Abstract

In this study we develop and test an econometric model for combining choice and preference ratings data collected from the same set of individuals. Choice data are modeled using a multinomial logit framework, while preference data are modeled using an ordered response equation. Individual heterogeneity is allowed for via random coefficients providing a link between the choice and ratings data. Parameters are estimated by Simulated Maximum Likelihood. An application of the model to consumer yoghurt choice in The Netherlands found that ratings based preference estimates differ significantly from choice based estimates, but the correlation between random coefficients driving the two is very strong.

1. Introduction

When modeling consumer preferences in the random utility framework a researcher has a number of econometric techniques available. With *revealed preference* (RP) data, i.e., actual consumer purchase data, the techniques are often determined by the nature of the available data. However, if *stated preference* (SP) data, which represent consumer decisions in hypothetical market situations, are to be collected, the researcher has the flexibility to choose which modeling approach to apply and to design choice experiments in line with this approach. In the marketing and transportation research literature, conjoint analysis is a frequently applied SP research technique, which encompasses analysis of three types of consumer preference data: ratings, rankings and choice data (e.g., Ben-Akiva et al. 1992, Bradley and Daley 1994, Haaijer et al. 1998, Louviere et al. 1993, Louviere 1994). The models used to estimate preferences for these data types range from OLS to ordered probit or ordered logit for ratings and multinomial probit or logit for the data on choices and rankings. Other SP methods of preference elicitation, more commonly found in the field of environmental economics, are contingent valuation (CV) methods that address individuals' willingness to pay (WTP) for certain environmental policy changes (e.g., Adamowizc et al. 1994, Carson et al. 1996). Again there are a number of different models that support estimation of preference models based on CV type response data which may be implemented depending on the type of data collected, for example, single-bounded, multiple-bounded and open-ended approaches to measuring WTP.

Although the approaches differ considerably, they are generally wielded

for the same purpose of eliciting consumer preferences, and, whilst methodology changes, for the same set of underlying preferences, utility estimates based on any of these models would ideally be statistically indistinguishable (after possible correction for task based biases). Therefore, if two differing types of data sets relating to the same consumers' preferences are available, an efficient use of the available data suggests that we should be able to estimate the same preference parameters from both sets simultaneously. Herein lies the concern of the current work: providing a model enabling estimation of the same consumers' utility functions from different types of stated preference data simultaneously if they are based on the same underlying utilities, or to analyze the differences in utilities between response modes if they occur. In particular, we examine two of the most commonly used SP responses: preference ratings and choice data.

Research interest in combining sources of preference data has recently increased (e.g., Hensher et al. 1999). There are various potential advantages to this, such as the opportunity to exploit the various strengths and weaknesses associated with each data type, and the possibility to test whether the decision processes underlying the data types are the same. If this hypothesis is rejected, a joint model can be used to analyse where partial differences between consumer utilities driving ratings and choice come from, and to trace question specific psychological factors that bias the utility indexes. Data pooling may also be required for implementation of new and more complex models recently developed in consumer research, such as models for examining the dynamic aspects of consumer processes, where panel data may be required (Louviere et al. 1999).

Furthermore, if different data sets arise from identical underlying

consumer utilities, joint estimation will provide more efficient results. Another goal of joint estimation therefore is an efficiency gain. If both ratings and choice data contain useful information on the underlying preferences of respondents, using both of them will help to get more accurate estimates of the parameters driving the utility function. Specifically, when comparing ratings and choice data, an advantage of ratings data is that it enables unbiased estimation of parameters at the individual level through the use of ordinary least squares. Disaggregate estimation is less desirable with choice data, as the most commonly used multinomial logit (MNL) model is biased for a small number of questions per respondent and estimates may even be infinite (Bunch and Batsell, 1989). Thus cost-reduction may also be achieved in data collection if fewer ratings than choice questions are required to get to the same level of statistical reliability, and if respondents find it easier to respond to additional ratings questions than additional choice questions.

The aim of this paper is to provide a model, consistent with random utility theory, for combining data on SP ratings and choice responses for the same individuals. In doing so, we do not treat the data sets as independent, but allow for correlation between the choices and the ratings of the same respondents. We model the ratings data with an ordered probit equation and the choice data via the multinomial logit model. Our modeling approach allows for heterogeneity across preferences in the population of consumers through random coefficients. This is advantageous because it allows for correlations between the choices and ratings for the same individual. According to random utility theory, the same consumer utility function should determine the outcomes in both data sets, and thus the

preference parameters driving choice and ratings data should be identical. This leads to testable restrictions on the parameters in the ratings and choice parts of the model.

We test the validity of this assertion, using data on yoghurt choices and ratings from a large consumer panel. We find that although consumers' preference ratings and choices are significantly correlated, there are significant differences in the standard deviations and some of the means of the random coefficients. Possible explanations for the observed differences drawn from the economic and psychological literatures are tested and discussed.

The remainder of the paper is organized as follows. Section 2 provides a brief review of the literature. Section 3 introduces the model. Data and results are given in Section 4. Some potential psychological and economic explanations for our findings are presented in Section 5. Section 6 concludes.

2. Literature review

Previous research on comparing SP ratings and choices has focused mainly on the predictive performance of models estimated on the different types of responses. In particular, Elrod et al. (1992) found that ratings and choice data generally perform equally well in terms of prediction at the aggregate level. The few studies examining the equivalence of the estimated preference parameters were predominantly done in the area of environmental economics. For example, Boxall et al. (1996) compared estimation results from choice data to those based on a CV WTP data set. They compare the welfare estimates based on the two data sets and find that the CV WTP estimate is over 20 times higher than the alternative SP

choice experiment suggests. The authors suggest the dramatic difference could be due to respondents misunderstanding the scenario, or a bias due to 'yea-saying', but believe that it is more probably a result of the respondents ignoring substitution possibilities in the CV questionnaire. Another study comparing data based on different elicitation methods is Cameron et al. (1999) who combine data arising from one RP choice, three SP choice tasks, one SP rating task and two WTP tasks, administered to seven independent samples. Their results indicate that once scale differences are allowed for, the hypothesis of equivalence of underlying utilities cannot be rejected across the choice and rating data sets, but do differ between the willingness to pay responses and the other responses. Likewise, Boyle et al. (1996) compare SP choice with WTP responses using three independent samples and find differences in scale between all data sets and differences in (relative) mean parameter estimates between two of their three data sets.

Other comparisons of preference elicitation methods have focused on the comparison between choice based models. A distinction can be made between papers that combine RP with RP or SP with SP (Morikawa 1989; Hensher and Bradley 1993, Swait et al. 1994) and those combining RP and SP (Louviere et al. 1993, Adamowicz et al. 1994, Bradley and Daley 1994). Both streams examine the hypothesis that consumer utilities underlying the pooled choice data sets are identical. The majority of these studies have found that after correcting for scale differences in error variance, the hypothesis of common preferences is not rejected.

In summary, the empirical evidence to date suggests that within a given response format, consumer utilities are mostly stable, but that there may be biases

associated with different survey response formats causing differences in response and/or utilities, especially between WTP and choice data responses. The difference between SP ratings and choices however, is not as well explored. Predictions on hold out consumer choice tasks based on SP ratings and choices do not seem to be seriously affected by response differences (Elrod et al. 1992). Also, after correcting for scale differences Cameron et al. (1999) could not reject the hypothesis of equal parameters underlying SP ratings and choice.

However, to date no econometric model has been proposed to combine and compare consumer ratings and choice data that allows for correlation between observations from the same individual. This limits the interpretation and testing of utility estimates based on SP ratings and choice, because individuals' responses to the two types of SP tasks cannot be integrated. It also limits possible efficiency gains both in terms of statistical estimation efficiency and in terms of data collection. Furthermore, developing insights into complex consumer behavior may require collection of multiple data types of the same individual in which case models allowing for individual responses to be correlated will be useful also.

3. Modeling consumer stated preference ratings and choice responses

In this section we present the econometric model to analyze consumers' SP ratings and choice data. We address issues of identification and scaling between models based on ratings and the choice data (cf. Swait and Louviere, 1993). For clarity of exposition, we first discuss the (more intuitive) model of consumer choice and then extend our model to include rating responses. We use the following notation:

i respondent (i=1,...,N); N is the total number of respondents

k attribute (k=1,…,K); K is the total number of attributes

s choice situation (s=1,…,S); S is the total number of choice situations

j alternative (j=0,1,..,J(s)); J(s) is the number of alternatives in choice situation s

J total number of different alternatives across all choice situations

$X_j = (x_{j1},...,x_{jK})'$ vector of attributes of alternative j; X_j does not include a constant.

3.1 Model for choice

Let the utility of alternative j for respondent i be given by:

(1) $U_{ij} = X_j'ß_i$ j=1,...,J

The vector of slope coefficients $ß_i=(ß_{i1},...,ß_{iK})'$ may vary across respondents. It reflects unobserved heterogeneity in the marginal utilities of the attributes.

Let alternative j=0 be the so called 'none'-option of not choosing any of the alternatives j. Its utility to respondent i is given by

(2) $U_{i0} = ß_{i0}$

The 'none'-option differs from the other alternatives in the sense that it does not have any attribute values. An equivalent way of modeling this utility would be to normalize the utility of the numeraire to 0, and add a respondent specific base level utility (which does not vary over attributes or alternatives) to the utility values of all the other alternatives.

The $ß_i$ and $ß_{i0}$ are treated as random coefficients, using the following specification:

(3) $ß_{ik} = b_k + u_{ik}, \quad k=0,...,K,$

(4) $u_i = (u_{i0},u_{i1},...,u_{iK}) \sim N(0,O)$

The unobserved characteristics of respondent i enter through u_{ik}. We assume that the u_{ik} are drawn from a (K+1)-variate normal distribution with mean zero. Note that $ß_i$ is respondent specific but not choice situation or alternative specific. It is thus assumed that the same $ß_i$ is used by respondent i in all choice situations. The parameters in the (K+1)×(K+1) matrix O are to be estimated. For computational convenience, we will assume that O is diagonal, so that only (K+1) standard deviations ($? _k$) need to be estimated. Since the random coefficients $ß_{i0}$ and $ß_i$ (or the u_{ik}) do not vary with choice situations or alternatives, and since they are independent across individuals, the correlation structure of choices across individuals, choice situations, and alternatives identifies the variances of the random coefficients.

In constructing a model for choice probabilities, we follow the usual multinomial choice framework in that choices are based upon the sum of utility values U_{ij} and errors $ε_{ijs}$:

(5) $U_{ijs}^* = U_{ij} + e_{ijs} \qquad j=0,\ldots,J(s),\ s=1,\ldots,S$

Respondent i chooses alternative c in choice situation s if and only if $U_{ics}^* \geq U_{ijs}^*$ for all j in the given choice situation. We assume that:

1. e_{ijs} is independent of exogenous variables (X) and random coefficients ($ß_i, ß_{i0}$).
2. $e_{ijs} \sim$ GEV(I), and
3. All e_{ijs} are independent of each other.

These assumptions imply that, conditional on the parameters $ß_{i0}$ and $ß_i$[1], we get the familiar multinomial logit choice probabilities:

(6) $P_{is}(c|ß_{i0},ß_i) = P(\text{i chooses alternative c in situation s}|\ ß_{i0},ß_i) = \dfrac{\exp(U_{ic})}{\sum_{j=0}^{J(s)} \exp(U_{ij})}$.

Here the summation is over the J(s)+1 alternatives in the given choice situation s (including the none-option). Moreover, for different choice situations, the choices of individual i are independent conditional on $ß_{i0},ß_i$. Thus the conditional probability for individual i with choice situations s = 1, ..., S, given $ß_{i0},ß_i$, to choose $J(i,1)$, ..., $J(i,S)$ is:

(7) $$LC_i(b_{i0}, b_i) = \prod_{s=1}^{S} P_{is}(J(i,s)|\ b_{i0}, b_i)$$

Normalization and identification

As usual, the scale of the utility function is normalized by a specific choice of the scale of e_{ijs}. This is the same as in a standard logit or multinomial logit model. The location parameters of the utility function ($ß_0$) are normalized by excluding the constant from X_j. As a consequence, all parameters determining the distribution

[1] Throughout, we also condition on the exogenous variables X, without mentioning this explicitly.

of the random coefficients are identified.[2]

3.2 Model for ratings

We refer to a task as a SP ratings task if an individual assigns a score on a scale (graphically or numerically) to a product, indicating the individual's preference for that product. SP ratings tasks differ from choice tasks in several respects. From the modeler's perspective, two important differences are that ratings responses are numerical or ordinal in nature, whereas choices are nominal, and that ratings are asked separately for each product, while choices often involve trade-offs between multiple products. To make the theoretical link between SP choices and ratings responses we assume that the ratings answer is based upon comparing the utility value of product j, (U_{ij}), to the utility of the numeraire (i.e., not buying the product) (U_{i0}). We will show below that this assumption is plausible given the wording of the ratings questions in our survey. Thus, we assume that an error free rating would be based upon $U_{ij} - U_{i0}$. Analogously to the error terms e_{ijs} in the choice model, we allow for a random error term, v_{ij}, and assume that the observed ratings are based upon

(8) $$U_{ij} - U_{i0} + v_{ij}$$

We assume that the error terms v_{ij} are mutually independent, independent of the exogenous variables, and independent of all other error terms in the model. Moreover, we assume they are all drawn from the same normal distribution[3] with

[2] It would also be possible to add alternative specific error terms which are independent across alternatives and individuals, but remain the same for a given individual and alternative across choice sets. In our empirical work, we included these effects, but found that they did not play a significant role.

3 Alternatively, a GEV(1) distribution could have been used which would have been somewhat more in line with the choice part of the model. In the literature on ratings, however, the normal distribution is more common. We do not expect any substantial differences for the results

mean 0 and variance s_v^2. The v_{ij} can be seen as evaluation errors on the ratings. Consumer heterogeneity enters through U_{ij}, i.e. through the random coefficients $ß_{i0}$ and $ß_i$. Correlation between choices and ratings comes in through these random coefficients. It therefore seems reasonable to assume that the v_{ij} are independent of the GEV(I) errors e_{ijs} in the choice evaluations.

Often, rating responses are grouped in classes, either due to a categorical response scale introduced by the researcher or by the respondents' natural tendency to prefer certain numbers over others (e.g., 10, 20, 30, etc). In what follows, we treat the observed ratings as an ordered categorical variable with R possible outcomes, say r = 1, ..., R. If the original ratings variable in the data is continuous, we first summarize it into a categorical variable before applying our model. We will come back to this below in discussing our data. We thus use an ordered response specification to model in which category the ratings are, similar to an ordered probit model. There is no reason why the scale of the utility function in the choice part (which is determined by normalizing the variance of the error terms in the choice part) would be the same as the scale of the ratings. Instead, it seems reasonable to allow for an unknown monotonic (possibly non-linear) transformation that transforms a utility index into a rating. This can be achieved in a flexible and simple way, by allowing for unknown bounds of the categories in the ordered response model.

To be precise, we assume that the ratings on a continuous scale underlying the categorical ratings are based upon the following unknown strictly increasing function g of the index in (8).

(9) $$R_{ij}^* = g(U_{ij}-U_{i0}+v_{ij})$$

We assume that g is the same for all respondents. As will be shown below, the assumption is needed to get the tractable ordered response model with fixed category thresholds. The assumption of fixed category thresholds is fairly standard in the ordered response models literature.

To transform the continuous (unobserved) variable R_{ij}^* into an observed categorical variable R_{ij} with R possible outcomes, we follow the same procedure as in a standard ordered response model. We partition the real line into R ordered categories, bounded by R-1 thresholds, and follow the standard assumption that these thresholds are common to all respondents. For notational convenience, the thresholds are denoted by $g(m_1)$, …., $g(m_{R-1})$. The link between R_{ij}^* and the observed categorical ratings, is now given by

(10) $R_{ij} = r$ if and only if $g(m_{r-1}) < R_{ij}^* \leq g(m_r)$ $(r=1, \ldots, R)$

Using (9) and the fact that g is strictly increasing, this can be rewritten as

(11) $R_{ij} = r$ if and only if $m_{r-1} < U_{ij}-U_{i0} + v_{ij} \leq m_r$

The thresholds $(-\infty = m_0 <) m_1 <... < m_{R-1} (< m_R = \infty)$ are unobserved parameters which can be estimated. Note that this procedure allows for an unknown strictly increasing transformation g, but g itself needs not to be estimated. This is the advantage of treating the ratings as an ordered categorical variable. Allowing for arbitrary values m_1, ..., m_{R-1} corresponds to using a flexible function g. To attain the same flexibility with a regression model for ratings observed on a continuous scale, it would be necessary to estimate g non-parametrically. We avoid this, and, instead, we only need the R-1 threshold values m_1, ..., m_{R-1}. These values are estimated as separate (ancillary) parameters.

Normalization and identification

If a model for the ratings only would be estimated, some normalization of scale and location would be necessary. One way to achieve this would be to fix U_{i0} and s_v^2 *a priori*. If, however, we simultaneously use the choice data (and use the same utility values in (8) as in (5)), the normalization is already imposed in the choice part of the model: the scale of U_{ij} is determined by the normalization of the variance of e_{ijs}. The constant term in the ratings corresponds to $ß_{i0}$ in the choice model, and is also identified (because the constant term is excluded from the other U_{ij}). In other words: there is no need for further normalization to identify the joint model for choice and ratings, and all the thresholds m_r (r = 1, …, R-1) can be estimated without imposing further restrictions.

3.3 Estimation and testing

In the joint estimation of the two parts of the model, using both choice and ratings data, the link between choice and ratings comes in through the random coefficients. For a given respondent, $ß_{i0}$ and $ß_i$ enter both the choice and the ratings. This distinguishes the estimation problem from the problem of estimating parameters using two or more independent samples, which is the more common situation in this literature (e.g., Boyle et al. 1996, Cameron et al. 1999).

We use smooth simulated maximum likelihood to estimate the model and to do inference. The likelihood is described below. A discussion of the estimation procedure and how its relation to standard estimation procedures is given in Appendix 1.

Likelihood contributions

Conditional on $ß_{i0}$ and $ß_i$, i.e. conditional on the U_{ij}, the probability that respondent i gives a specific series of M categorical ratings, can be written as the product of univariate normal probabilities (as in an ordered probit model). Moreover, conditional on $ß_{i0}$ and $ß_i$ the ratings are independent of the choices, so the conditional probabilities of the observed categorical ratings and the observed choices, given $ß_{i0}$ and $ß_i$, are the product of choice and ratings contributions. Conditional on $ß_{i0}$ and $ß_i$, the likelihood contribution of a given respondent is therefore a product of univariate normal probabilities (ratings) and MNL probabilities (choice part). The unconditional likelihood is the expected value of the conditional contribution, with the expectation taken over the (joint) density of $ß_{i0}$ and $ß_i$, a (K+1)-dimensional integral for which no analytical expression can be given.

A test for preference stability

There are several strategies for constructing tests of whether ratings and choice are indeed driven by the same preferences. A test which does not require a specific alternative model would be a Hausman test (see Hausman 1978), comparing the estimates using ratings as well as choice data (efficient under the null) with the estimates based upon the choice data only (inefficient under the null, consistent under the alternative). A problem with the standard way of performing the test is that the estimated difference of the two covariance matrices is not positive definite - although it should asymptotically be positive definite under the null. Moreover, the power of this test could be limited. Since we do have particular alternatives in

mind here, a more natural way to go is to formulate a more general model which nests the joint model introduced above but has separate utility indexes underlying ratings and choices, and perform a Likelihood Ratio (LR) or a Lagrange multiplier test. We will use the LR test, since the estimates of the more general model are of some interest by themselves, possibly indicating why the joint model is rejected.

A more general model can be formulated as follows. The natural generalization of the joint model is that the ratings are not generated by (8) but by a separate utility index

(12) $V_{ij} = -\alpha_{i0} + X_j'\alpha_i$

(13) $\alpha_{ik} = a_k + \eta_{ik}, \quad k=0,...,K.$

Similar assumptions are made on the distribution of $\eta_i = (\eta_{i0},\eta_{i1},...,\eta_{iK})$ as on u_i (but with potentially different parameters). It seems reasonable to allow for an arbitrary correlation coefficient between η_i and u_i. A parsimonious way to achieve this, is the following specification of η_{ik}:

(14) $\eta_{ik} = \lambda_k[\rho u_{ik} + (1-\rho)w_{ik}],$

with $w_{ik} \sim N(0,1)$, mutually independent and independent of other error terms and of exogenous variables. If $\rho=0$ (14) implies that random coefficients in ratings and choice are independent, and the model partitions into independent models for ratings and choice. Without restrictions on the parameters across the two parts of the model, ML (or simulated ML) estimates for this model with $\lambda=0$ will be the same as ML estimates for separate ratings and choice models. If $\rho=1$, the η_{ik} are perfectly correlated to the u_{ik}, though they still may have different variances, and the random coefficients may still have different means and variances.

In the general model, two constraints have to be imposed on the ratings part of the model, since scale and location of this part of the model are not identified without imposing restrictions across ratings and choice part. We set $\sigma_v=1$ and $a_0=b_0$. The joint model discussed above results if we impose the restrictions

(15) $?=1$, $a_k = b_k$ $(k=0,\ldots,K)$, $?_k = 1$ $(k=0,\ldots,K)$.

These are 1+2(K+1) restrictions, but this is partly compensated by the two restrictions needed to identify the general model. Thus the Likelihood Ratio test statistic will, under the null that the joint model is valid, asymptotically follow a chi squared distribution with 2(K+1)-1 degrees of freedom.

4. Empirical analysis

4.1 Data

The survey analyzed in this study was concerned with the evaluation of hypothetical yoghurt products, a commonly consumed commodity in the market that was studied. Data were collected using a survey distributed to respondents participating in the CentERdata consumer panel. This panel consists of consumers from throughout The Netherlands and is administered by Tilburg University since 1998. Respondents were screened for regular yoghurt consumption, and of the 977 respondents surveyed, 909 remained after incomplete and incorrect responses were removed.

In the survey, respondents were asked to imagine having lunch in a cafeteria and having to decide whether or not to purchase a 200ml container of yoghurt with

their meal. The attributes considered in the survey and their levels are summarized in Table 1. Attributes and their levels were selected after a thorough examination of yoghurt products in local supermarkets, and discussions with regular yoghurt consumers. A total of 7 attributes, each presented at 2 levels, were used in the presentation of products: 3 continuous variables (price, fruit content, fat content) and 4 binary variables (biological cultures, artificial flavouring, creamy taste, recyclable packaging).

- INSERT TABLE 1 ABOUT HERE -

To control for the possible effect of attributes not included in the study, respondents were instructed to assume that the yoghurts were identical with respect to all characteristics not presented in the survey and were available in their favorite flavour. Furthermore, they were advised to assume there were no other yoghurts available in the cafeteria when considering each separate question.

Statistical design methods, following Louviere and Woodworth (1983), were used to construct product profiles and choice sets in which attributes were orthogonal. To calibrate the attribute levels a small survey was conducted from which preliminary marginal utility contributions were estimated for each attribute. Using this information, the levels of the continuous attributes were adjusted so that the predicted change in utility between the two levels considered was approximately equal to the average change in marginal utility associated with the binary attributes. Maintaining utility balance across attributes is important for improving the efficiency of statistical designs (Huber and Zwerina, 1996).

Each participant in the survey was first asked to rate eight yoghurt products and then to complete a series of eight choice questions. Half of the respondents were also given eight hold-out choice questions that were used for further model validity testing (see section 4.3). The design of the rating and choice tasks is as follows.

Ratings task

With seven attributes each described at two levels, $2^7 = 128$ distinct product profiles can be created, which if all combined in the same survey questionnaire would result in an orthogonal array of attribute levels. The fact that the total number of possible combinations increases so rapidly, has led to increased use of fractional factorial designs (see Green, 1974), which greatly reduce the number of product profiles to be presented whilst maintaining orthogonality between the main effects of the attributes. The use of such orthogonal arrays presents one of the major advantages of SP data over RP data, as the latter is often found to exhibit collinearity between attributes, hampering identification of the marginal contribution of different attributes. Using a 1/16 fraction main effects design produced eight mutually orthogonal product profiles.

All subjects were presented with each of the eight product profiles and asked to separately indicate for each product, on a scale of 0 - 100, the probability that they would purchase the yoghurt if there were *no other yoghurts* available in the cafeteria. Probability ratings tend to have a good rationale for predicting choice compared to other forms of ratings data (Elrod et al. 1992, Wittink and Cattin 1989). Moreover, phrasing the question as a probability of purchase makes

it reasonable to assume that the rating scores are based upon comparing the utility of each alternative with the utility of the 'none' option of not buying any yoghurt product. This assumption is made in the modeling section. The same 'none' option is also incorporated in the SP choice sets (see below).

As explained in the previous section, we do not use the exact ratings on the continuous scale 0 – 100, but first transform them into categorical levels. Since the frequencies in the data show clear peaks at multiples of 10, we used eleven categories: 1 if the rating is less than 10, 2 if the rating is greater than or equal to 10 and less than 20, and so on with category 11 representing ratings of 100.

Choice task

After the eight ratings questions, each respondent answered eight choice questions. In each of these, respondents were asked to choose one option from a hypothetical choice set including yoghurt products and the 'none' option. The choice sets contained two products, which were again described by bundles of the attributes introduced above. One option in each choice question was constructed based upon the same eight profiles that were used to construct the ratings questions. The other option was its socalled 'foldover' profile, which in the case of binary attributes is the product with the exact opposite attribute levels. This approach guaranteed orthogonality within and between the two yoghurt options in the different choice sets. Moreover, having a constant reference alternative (the 'none' option) in each choice set guarantees that the choice sets exhibit orthogonality not only in attributes but also in *attribute differences*. Orthogonality

in attribute differences is statistically more important than orthogonality in attribute levels for identification of main effects for 'difference-in-utility' models such as the MNL model (Louviere, 1994; Louviere and Woodworth, 1983).

4.2 Estimation results

Table 2 presents the results of the joint model estimated on the ratings questions and the eight choices for each respondent.[4] The means of the random coefficients all have the expected sign and are strongly significant. The confidence intervals for the standard deviations of the random coefficients never contain the value zero, indicating significant heterogeneity in preferences between respondents.

- INSERT TABLE 2 ABOUT HERE -

To test the joint model formally, we also estimated the more general model using (12) to (14). The estimation results are presented in table 3. As in table 2, all parameters have the correct sign and are significant at the 5% level. Estimated means of the random coefficients for ratings and choices are of the same order of relative magnitude, with some notable exceptions. In particular, the price effect in the ratings estimates is about 20% larger than in the choice estimates, suggesting that ratings are more sensitive to price than choice. Furthermore, 'biological cultures' and 'recyclable packaging' also are relatively larger in the ratings estimates. Most of the estimated standard deviations for ratings and choice parameters are similar in magnitude.

The estimated value of *?* was 0.937 with standard error 0.023. Thus *l* is

[4] All results are based upon T=40 draws in the simulated ML procedure for each respondent.

significantly different from zero as well as from 1. (The latter result is also obtained using a Likelihood Ratio test.) The model with l=0 is the same as the combination of two separate independent models for ratings and choice. Thus the result that l is significantly different from 0 implies that ratings and choice data cannot be treated as independent samples. The result that the estimate of l is close to 1 implies that knowledge of a specific respondent's utility function based on their ratings, would also be informative about their choice probabilities. Although the coefficients differ in mean and dispersion, they are strongly correlated. Thus, combining the two data sources can be expected to provide a more stable basis for segmenting consumer populations in terms of their preferences.

- INSERT TABLE 3 ABOUT HERE -

Considering the small standard errors, the difference in parameters between ratings and choice can be expected to be significant, suggesting that the joint model will statistically be rejected against the more flexible model. To test this observation formally, a Likelihood Ratio test was conducted comparing the joint model to the general model. This test rejected the null hypothesis that ratings and choice are based upon the same utility indices. Some further tests of hybrid models allowing for more flexibility in the joint model were also conducted. All hybrid models were rejected against the general model. The log-likelihood values of these models and the successive differences are reported in table 4.

- INSERT TABLE 4 ABOUT HERE -

The fact that ratings and choices are far from independent can also be confirmed in another way. Separate estimations of the choice model and the ratings model (after adding an appropriate normalisation to the latter) give log-likelihood values summing up to -20864.8. This sum is the log-likelihood of a combined model that imposes independence of random coefficients in ratings and choice (?=0 in (13)). According to a Likelihood Ratio test, this model is rejected against the general model. It is interesting to note that the likelihood of the model imposing independence is also much lower than the likelihood of the much more parsimonious joint model. Although the two models are non-nested so that a standard Likelihood Ratio test cannot be performed, this shows that the joint model performs much better than a model imposing independence (although even the joint model is rejected against the general model with dependence).

4.3 Predictive tests on hold out choices

Although an efficiency gain is obtained in estimating the parameters of the choice model by using the ratings data, the question seems justified whether using the ratings data affects predictions of consumer choice. And if so, if the more parsimonious joint model or even a choice only model might not predict equally well as the flexible general model. We address this question by looking at some predictions for three alternative choice situations. For this purpose, we use the hold out choice questions answered by the respondents.

The difference between the hold-out groups was only in terms of the number of alternatives that were presented in each choice set. For hold-out group

1, the new choice sets are of the same type as the old ones (two products and the none option). In hold-out group 2, respondents evaluated four alternatives (plus the none option), none of which was dominated by one of the others. Respondents in hold-out group 3 evaluated choice sets with six non-dominated alternatives and the none option.

Predictions for the joint and general model are generated in the following way. For each respondent, 20 values of the random coefficients are generated using the estimated parameters. In case of the general model only the choice parameter estimates were used. Based upon these coefficients, the utility values of each of the alternatives in each of the eight new choice situations are predicted. This gives choice probabilities for all the alternatives[5], and we have computed the averages of these probabilities in each hold-out group.

The predicted shares are compared to the actual shares in the hold out data. We have summarized the results in terms of mean absolute deviation, where the mean is taken over the alternatives in each choice set and over the eight choice sets. This is done for the parameter estimates of the choice only model, the joint model and the general model. Results are given in table 5.

All models performed quite well. It can be seen that all three models performed very similarly in terms of predictive accuracy, with a small advantage for the joint model. The improvement in predictive performance of the joint model over the choice only model was only small. This was especially so for the two hold-out groups where more alternatives per choice set were evaluated than in the original choice sets.

[5] The none-option is treated in the same way as the other alternatives.

- INSERT TABLE 5 ABOUT HERE -

5. Discussion

Various theoretical approaches can be taken to explain the observed differences in ratings and choice estimates. In reviewing the relevant literature, the more psychologically oriented set of potential explanations can be distinguished from the more economically oriented set. Olsen et al. (1995) give a good review of the former, while Carson et al. (1999) review the latter. The different explanations that are suggested are now briefly reviewed and tested on our findings.

5.1 Psychological explanations

A first possible explanation found within the psychological literature is the *prominence effect* (e.g., Tversky et al. 1988). This effect occurs if the most important piece of information in the description of an alternative receives greater weight in a choice task than in a judgment task such as a rating. The underlying explanation is that in judgment tasks respondents tend to use more compensatory evaluation processes than in choice, taking into account more aspects of the alternatives. As a consequence, choice based estimates would have higher values for the most important attributes. Our results may perhaps be explained in part by this effect. After correcting for coding differences (multiplying with ranges for each attribute), the most important attributes in terms of utility both in the ratings and choice responses were fat content and artificial flavoring (see table 6). Although the difference was not large, the relative value of these two parameters

compared to all other parameters except for fruit was higher in the choice estimates than in the ratings estimates, providing some support for the prominence effect.

- INSERT TABLE 6 ABOUT HERE -

A related explanation that has been suggested is that given that judgment tasks lead to more compensatory evaluations (Billings and Scherer 1988, Einhorn et al. 1979) more attributes should be of importance and/or significance in the ratings estimates, while fewer parameters should be so in the choice estimates. This effect occurred only to a minor degree in our findings. All attributes were significant in the estimates for both response types. Also, the relative size of the attributes was largely similar over response modes, possibly with the exception of recyclable packaging, which was relatively more important in the ratings responses (see table 6).

A second possible psychological explanation can be found in the *compatibility effect* (e.g., Montgomery et al. 1994). This effect indicates that product information that is presented in a format that is more similar to the response format will receive greater weight in the evaluations. The underlying explanation for the effect is that cognitive switching costs are lower between similar types of information, making it easier to include information that matches with the response task in the evaluation. On the basis of this effect one would expect the attributes price, fruit and fat content to have a greater relative importance in the ratings estimates, while the (dichotomous) other attributes

should have greater importance in the choice estimates. This effect is rejected by our results (see table 6).

5.2 Economic explanations

The economic literature in this area stresses the potential for *strategic behavior* on the part of the respondent (Carson et al. 1999). It is assumed that the respondents act rationally in choosing which information they wish to provide to manufacturers. Therefore, different response formats and different assumptions that consumers may make with respect to manufacturers' intentions are expected to lead to different strategic incentives for respondents.

In our study, the two response formats have the following relevant aspects. In the ratings task, consumers are asked to evaluate an alternative over the option of not buying. In the choice task, a comparison is made between two alternatives, while the option of not buying is included also. In both cases, the likely assumptions with respect to the manufacturer's intentions that consumers may make are that on the basis of the consumer's responses the manufacturer may: 1. Decide on the optimal price and promotions level to set for its yoghurt products, and 2. Decide on whether or not to introduce a new yoghurt product in the market, and if so, which new products to introduce.

In response, the rational consumer will choose an answering strategy that strategically speaking should lead to lower manufacturer pricing and more new product introductions, especially introduction of products that are liked by the consumer. This behavior is rational because it reduces consumer costs and increases the number of consumer choice options at no additional cost.

To achieve this type of desirable manufacturer response, the strategically optimal consumer response strategy differs for the two response formats. In the ratings responses, consumers should indicate a relatively low willingness to pay for existing products and a relatively high willingness to pay for new products. Note that this strategy is not in line with revealing the consumer's true preferences for different attributes. In particular, the observed price sensitivity can be expected to be higher than the consumer's true price sensitivity (leading to lower manufacturer pricing), and the consumer's utility for new product features can be expected to be higher than the consumer's true utility (leading to more new product introductions). In the case of choice responses, the strategically optimal consumer response is more aligned with responding according to their actual preference. If in comparing the two alternatives, the consumer makes the assumption that only one of the alternatives will be introduced in the market, it is in the consumer's interest that only his or her most preferred product is introduced. Therefore, in the trade off between the two products it is in the consumer's interest to reveal their true preference and price sensitivity. In the comparison with none, similar considerations exist as in the ratings task, so that even the choice based estimates may not be fully in line with the consumer preferences.

Based on the differences in strategic incentives between the two response formats, one would expect to find higher price sensitivity in the ratings task and higher utility estimates for possible new attributes in the ratings task. Because the attributes biological cultures and recyclable packaging currently are not offered in most cafeterias, consumers could regard these as possible product innovations.

Thus it can be expected that these attributes should receive relatively higher utility estimates in the ratings parameters. These expectations are supported by our results. The relative size of the price parameter and the estimate for recyclable packaging are higher for the ratings responses, thus providing support for the economic explanation. Because the parameter for biological cultures was used as a minimum benchmark for both response types, its relative size could not be established.

6. Conclusion

We have developed a model to combine and compare consumer utility estimates based on stated preference ratings and choice responses. The modeling approach combined two components: a random coefficients ordered probit to model consumers rating responses and a random coefficients logit to model consumers' choices. Correlation between the two components was introduced through the random coefficients in the model. An empirical application of the proposed model illustrated its flexibility in comparing and combining parameter estimates based on consumer ratings and choice data.

In our empirical results we found significant differences between ratings based and choice based utility estimates. In particular, respondents were relatively more price sensitive in the ratings tasks as well as more positive about possible new product extensions (i.e., recyclable packaging). These observed effects were in line with possible strategic behavior by consumers in responding to the survey questions. Some support was found also for the prominence effect indicating that the most important attribute received greater weight in the choice task. No

support was found for the compatibility effect.

Despite these differences in parameters it was found that the predictive ability of the different models was very similar. This finding may seem surprising, but is in line with earlier results by Dawes (1979) who showed that linear models perform very well in predicting the outcome of choice tasks even if the linear models are only directionally correct and the parameter values have incorrect values. Empirical results by Elrod et al. (1992) also illustrate a similar predictive ability of different model specifications based on consumer ratings and choice responses, further supporting the view that aggregate predictions are robust over utility measurement approaches.

Given that strategic response behavior can explain part of the observed differences between ratings and choices in our estimates and the fact that choice tasks are less prone to strategic respondent behavior, the results suggest that choice responses may be more suitable if one wishes to understand consumer preference structures. Carefully designed choice experiments can be used to avoid potential biases due to strategic behavior. Further research in this area could explore consumers' inclination to respond strategically under different conditions (e.g., by changing the context presented in the study). Based on our findings future research also may address the possible value of combining ratings and choice responses in consumer segmentation research. For example, segmentation may be more successful if one takes into account the correlation in individuals' ratings and choice responses. The cost efficiency of collecting these two types of responses simultaneously may also be studied, trading off the costs of additional data collection per respondent against the costs of collecting data from more

respondents. If the prediction of market shares is the objective however, collecting data in one response format may be equally suitable.

References

Adamowicz, W. L., Louviere, J. J., and Williams, M. (1994), "Combining Stated and Revealed Preference Methods for Valuing Environmental Amenities," *Journal of Environmental Economics and Management*, 26, 271-276.

Ben-Akiva, M., Morikawa T., and Shiroishi, F. (1992), "Analysis of the Reliability of Preference Ranking Data," *Journal Business of Research*, 24, 149-164.

Billings, R. S., and Scherer, L. L. (1988), "The Effects of Response Mode and Importance on Decision-Making Strategies: Judgment Versus Choice," *Organizational Behavior and Human Decision Making Processes*, 41, 1-19.

Boxall, P. C., Adamowicz, W. L., Swait, J., Williams, M., and Louviere, J. J. (1996), "A Comparison of Stated Preference Methods for Environmental Valuation," *Ecological Economics*, 18, 243-253.

Boyle, K. J., Johnson, F. R., McCollum, D. W., Desvousges, W. H., Dunford, R. W., and Hudson, S. P. (1996), "Valuing Public Goods: Discrete Versus Contingent Valuation Responses," *Land Economics*, 72, 381-396.

Bradley, M., and Daley, A. (1994), "Use of the Logit Scaling Approach to Test Rank-Order and Fatigue Effects in Stated Preference Data," *Transportation*, 21, 167-184.

Bunch, D., and Batsell, R. (1989), "A Monte Carlo Comparison of Estimators for the Multinomial Logit Model," *Journal of Marketing Research*, 26, 56-68.

Cameron, T. A., Poe, G. L., Ethier, R. G., and Schulze, W. D. (1999), "Alternative Nonmarket Value-Elicitation Methods: Are Revealed and Stated Preferences the Same?," Working Paper, Department of

Economics, UCLA, December 1999.

Carson, R., Flores, N., Martin K., and Wright, J. (1996), "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates of Quasi-Public goods," *Land Economics*, 72, 80-99.

Carson, R. T., Groves, T., and Machina, M., (1999), "Incentive and Informational Properties of Preference Questions," plenary adress European Association of Resource and Environmental Economists, Oslo, Norway, June 1999.

Dawes, R.M. (1979), "The Robust Beauty of Improper Linear Models in Decision Making," *American Psychologist*, 34, 571-582.

Einhorn, H. J., Kleinmuntz, D. N., and Kleinmuntz, B. (1979), "Linear regression and process tracing models of judgment," *Psychological Review*, 86, 465-485.

Elrod, T., Louviere J. J., and Davey, K. S., (1992), "An Empirical Comparison of Ratings-Based and Choice-Based Conjoint Models," *Journal of Marketing Research*, 29, 368-377.

Green, P. E., (1974), "On the Design of Choice Experiments Involving Multifactor Alternatives," *Journal of Consumer Research*, 1, 61-68.

Haaijer, R., Wedel, M., Vriens, M., and Wansbeek, T. (1998), "Utility Covariances and Context Effects in Conjoint MNP Models," *Marketing Science*, 17, 236-252.

Hajivassiliou, V. A., and Ruud, P. A. (1994), "Classical Estimation Methods for LDV Models Using Simulation," in *Handbook of Econometrics*, eds. R. F. Engle, and D. L. McFadden, Amsterdam: North Holland, 2384-2438.

Hausman, J., (1978), "Specification tests in econometrics," *Econometrica* 46,

1251-1271.

Hensher, D. A., and Bradley, M. (1993), "Using Stated Preference Response Data to Enrich Revealed Preference Discrete Choice Models," *Marketing Letters*, 4, 39-152.

Hensher, D. A., Louviere J. J., and Swait, J. (1999), "Combining Sources of Preference Data," *Journal of Econometrics*, 89, 197-221.

Huber, J., and Zwerina, K. (1996), "The Importance of Utility Balance in Efficient Choice Designs," *Journal of Marketing Research*, 33, 307-317.

Louviere, J. J. (1995), "Relating Stated Preferences Measures and Models to Choices in real Markets: Calibration of CV Responses," in *The Contingent Valuation of Environmental Resources*, eds. D. J. Bjornstad, and J. R. Kahn, Brookfield: Edward Elgar, 167-188.

Louviere, J. J. (1994), "Conjoint Analysis," in: Bagozzi, (ed.), *Advanced Methods of Marketing Research*, Blackwell Publishers, Cambridge, MA, 223-259.

Louviere, J. J., Fox, M. F., and Moore, W. L. (1993), "Cross-Task Validity Comparisons of Stated Preference Choice Models," *Marketing Letters*, 4, 205-213.

Louviere, J. J., Meyer, R. J., Bunch, D. S., Carson, R., Dellaert, B. G. C., Hanemann, W. M., Hensher, D., and Irwin, J. (1999), "Combining Sources of Preference Data for Modeling Complex Decision Processes," *Marketing Letters*, 10, 205-217.

Louviere, J. J., and Woodworth, G. (1983), "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data," *Journal of Marketing Research*, 20, 350-367.

Montgomery, H., Selart, M., Garling T., and Lindberg E., (1988), "Adaptive Strategy Selection in Decision Making," *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 534-552.

Morikawa, T. (1989), "Incorporating Stated Preference Data in Travel Demand Analysis," Ph.D. Dissertation, Dept. of Civil Engineering, MIT.

Olsen, G. D., Swait, J., Johnson, R. D., and Louviere, J. J. (1995), "Response Mode Influences on Attribute Weight and Predictive Ability When Linear Models are Not Certain to be Robust," Working Paper, Faculty of Management, University of Calgary.

Swait, J. and Louviere, J. J. (1993), "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models," *Journal of Marketing Research*, 30, 305-314.

Swait, J., Louviere, J. J., and Williams, M. (1994), "A Sequential Approach to Exploiting the Combined Strengths of SP and RP Data: Application to Freight Shipper Choice," *Transportation*, 21, 135-152.

Tversky, A., Sattath S., and Slovic, P. (1988), "Contingent Weighting in Judgment and Choice," *Psychological Review*, 95, 371-384.

Wittink, D. R., and Cattin, P. (1989), "Commercial Use of Conjoint Analysis: An Update," *Journal of Marketing*, 53, 91-96.

Appendix 1 - Smooth simulated Maximum Likelihood

To estimate the joint model[6] by simulated ML, the multi-dimensional integral in the unconditional likelihood is approximated by a simulated mean. This simulated mean is based upon draws of standard normal error terms which can be transformed into $ß_{i0}$ and $ß_i$. Let T denote the number of independent draws of all random variables that will be used per individual. T has to be chosen prior to estimation. Smooth simulated ML is then based upon the following steps.

1. Before starting the ML algorithm, draw (K+1)NT independent standard normal variables ζ_{ikt}
2. During a specific ML iteration, for given values of the parameters, the means and variances of $ß_{i0}$ and $ß_i$ are given by b_k, and $?_k^2$ (k=0,..,K; i=1,…,N). Now set $\beta_{ikt} = b_k + ?_k\zeta_{ikt}$. Thus the β_{ikt} can be seen as independent draws from $N(b_{ik},?_k^2)$, the correct distribution of the random variables $ß_{i0}$ and $ß_i$ which should be drawn). Stack them into (K+1)N vectors of length T: $\beta_{il}=(\beta_{i0l},...,\beta_{iKt})'$.
3. Instead of maximizing $S_i \log L_i$, maximize $S_i \log LS_i$, where: $LS_i = 1/T\sum_{t=1}^{T} L_i(\beta_{it})$.

Thus the expected value is replaced by a simulated sample mean of T draws. The Law of Large Numbers implies that for large T, LS_i will approximate L_i.

It can be shown that this procedure is asymptotically equivalent to ML provided that $T\rightarrow\infty$ fast enough (e.g., Hajivassiliou and Ruud 1994). This implies that standard ways of obtaining ML estimates, standard errors, etc. can be used. The approximated likelihood $S_i \log LS_i$ can be treated as the real likelihood. Since the e_{ijs} in eq. 5 and the v_{ij} in eq. 8 are not simulated, the simulated likelihood function

[6] The other models can be estimated in a similar way.

is a smooth (differentiable) function of the parameters to be estimated. This has several advantages over some of the early, non-smooth, simulated maximum likelihood methods (See Hajivassiliou and Ruud, 1994).

Table 1. Attributes and levels used in the experiment

Attribute	Description of levels	Coding in estimation
Price	NLG 1.90	1.9
	NLG 1.50	1.5
Fruit content	10% fruit	10
	5% fruit	5
Biological cultures	Contains biological cultures	1
	Contains no biological cultures	0
Artificial flavouring	Contains artificial flavouring	1
	Contains no artificial flavoring (all natural)	0
Creamy taste	Creamy taste	1
	Regular taste	0
Fat content	0.5% fat content	0.5
	3.5% fat content	3.5
Recyclable packaging	Yoghurt container is recyclable	1
	Yoghurt container not recyclable	0

Table 2. Estimation results: Joint model

Parameter	*Estimate*	*Standard error*
Mean coefficients		
($ß_{i0}$) – None Constant	-3.037	0.122
($ß_{i1}$) – Price	-1.285	0.064
($ß_{i2}$) – Fruit	0.141	0.005
($ß_{i3}$) – Biological Cultures	0.412	0.025
($ß_{i4}$) – Artificial Flavoring	-0.793	0.032
($ß_{i5}$) – Creamy Taste	0.476	0.025
($ß_{i6}$) - Fat Content	-0.355	0.011
($ß_{i7}$) – Recyclable Packaging	0.564	0.026
Standard deviations of Random coefficients		
(? $_0$) – None Constant	1.053	0.035
(? $_1$) – Price	0.473	0.018
(? $_2$) – Fruit	0.076	0.004
(? $_3$) – Biological Cultures	0.144	0.029
(? $_4$) – Artificial Flavoring	0.727	0.030
(? $_5$) – Creamy Taste	0.418	0.029
(? $_6$) - Fat Content	0.373	0.011
(? $_7$) – Recyclable Packaging	0.071	0.035
Category Thresholds		
m_1	-1.243	0.061
m_2	-0.579	0.048
m_3	0.122	0.037
m_4	0.592	0.035
m_5	1.124	0.037
m_6	2.106	0.048
m_7	2.725	0.060
m_8	3.625	0.079
m_9	4.535	0.101
m_{10}	5.442	0.126
Ratings error standard deviation (s_v)	1.540	0.042

Table 3. Estimation results: General model

	Choice part		*Ratings part*[7]	
Parameter	*Estimate*	*Standard error*	*Estimate*	*Standard error*
Mean coefficients				
($ß_{i0}$) – None Constant	-3.248	0.167	-	-
($ß_{i1}$) – Price	-1.186	0.091	-1.402	0.123
($ß_{i2}$) – Fruit	0.139	0.008	0.133	0.009
($ß_{i3}$) – Biological Cultures	0.360	0.042	0.477	0.045
($ß_{i4}$) – Artificial Flavoring	-0.870	0.044	-0.752	0.051
($ß_{i5}$) – Creamy Taste	0.429	0.038	0.532	0.044
($ß_{i6}$) – Fat Content	-0.403	0.015	-0.331	0.016
($ß_{i7}$) – Recyclable Packaging	0.478	0.042	0.694	0.047
Standard deviations of Random coefficients				
(? $_0$) – None Constant	1.539	0.075	1.246	0.025
(? $_1$) – Price	0.388	0.032	0.238	0.014
(? $_2$) – Fruit	0.096	0.007	0.077	0.004
(? $_3$) – Biological Cultures	0.053	0.059	0.132	0.043
(? $_4$) – Artificial Flavoring	0.767	0.057	0.707	0.042
(? $_5$) – Creamy Taste	0.452	0.048	0.309	0.041
(? $_6$) - Fat Content	0.392	0.019	0.377	0.011
(? $_7$) – Recyclable Packaging	0.003	0.055	0.114	0.046
Category Thresholds				
m_1			-0.935	0.321
m_2			-0.280	0.320
m_3			0.411	0.320
m_4			0.874	0.320
m_5			1.399	0.319
m_6			2.365	0.318
m_7			2.974	0.317
m_8			3.859	0.316
m_9			4.761	0.314
m_{10}			5.659	0.313
l	0.937	0.023		

7 For normalization purposes, β_{i0} in the ratings part of the model is set equal to β_{i0} from the choice part.

Table 4. Likelihood ratio test results

Model specification	*Likelihood*	*Difference with previous model*	*d.f. difference with previous model*
Joint model	-20607.4	-	-
Standard deviations differ	-20563.2	44.2*	8
Standard deviations and price parameter differ	-20559.4	3.8*	1
Standard deviations and price, biological cultures and recyclable packaging differ	-20544.8	14.6*	2
Standard deviations and all coefficients differ	-20532.6	12.2*	4
Standard deviations and all coefficients differ and l is estimated	-20530.6	2.0*	1
Independent models for ratings and choice	-20864.8	n.a.	

* significantly different from the previous (more parsimonious) model at 95% confidence level

Table 5. Mean absolute deviations from actual choice shares

Hold-out group	*Choice only model*	*Joint model*	*General model*
1 identical choices (n = 147)	0.092	0.071	0.077
2 four alternatives (n = 164)	0.050	0.047	0.051
3 six alternatives (n = 153)	0.044	0.044	0.045

Table 6. Comparison of ratings and choice based estimates

	Estimate corrected for coding differences		Importance ranking		Relative size*	
Parameters	Choice part	Ratings part	Choice part	Ratings part	Choice part	Ratings part
Price	-0.474	-0.561	4	5	0.13	0.16
Fruit	0.695	0.665	3	4	0.39	0.36
Biological cultures	0.360	0.477	7	7	0.00	0.00
Artificial flavoring	-0.870	-0.752	2	2	0.60	0.53
Creamy Taste	0.439	0.532	5	6	0.08	0.11
Fat content	-1.209	-0.993	1	1	1.00	1.00
Recyclable packaging	0.403	0.694	6	3	0.05	0.42

* Relative size is calculated as $R = \frac{|b_k| - |b_{\min}|}{|b_{\max}| - |b_{\min}|}$, where b_k is the relevant parameter and $b_{\min}$ and $b_{\max}$ are the parameters with the lowest and highest absolute value respectively (all corrected for coding differences).