Using Attribute Importance Rankings within Discrete Choice Experiments: An Application to Valuing Bread Attributes.

Kelvin Balcombe, Department of Agricultural and Food Economics, University of Reading

Michael Bitzios, Kent Business School, University of Kent

Iain Fraser*, School of Economics, University of Kent

and

School of Economics, La Trobe University

Janet Haddock-Fraser, Faculty of Social Science, Canterbury Christchurch University

Contributed paper prepared for presentation at the 57th AARES Annual Conference, Sydney, New South Wales, 5th-8th February, 2013

AARES
AUSTRALIAN AGRICULTURAL & RESOURCE ECONOMICS SOCIETY

Using Attribute Importance Rankings within Discrete Choice Experiments: An Application to Valuing Bread Attributes.

Kelvin Balcombe, Department of Agricultural and Food Economics, University of Reading

Michael Bitzios, Kent Business School, University of Kent

Iain Fraser*, School of Economics, University of Kent

and

School of Economics, La Trobe University

Janet Haddock-Fraser, Faculty of Social Science, Canterbury Christchurch University

***Address for Correspondence:**

School of Economics, University of Kent, Canterbury, Kent, CT2 7NP, UK

Tel: +44 (0)1227 823513

Fax: +44 (0)1227 827850

Email: i.m.fraser@kent.ac.uk

Abstract

We present a new Bayesian econometric specification for a hypothetical Discrete Choice Experiment (DCE) incorporating respondent ranking information about attribute importance. Our results indicate that a DCE debriefing question that asks respondents to rank the importance of attributes helps to explain the resulting choices. We also examine how mode of survey delivery (online and mail) impacts model performance, finding that results are not substantively affected by the mode of survey delivery. We conclude that the ranking data is a complementary source of information about respondent utility functions within hypothetical DCEs.

Key Words: Attribute Importance Rankings, Discrete Choice Experiment, Survey Mode

JEL Classification: C11, C25, L66

Introduction

There is a rapidly growing literature that examines how respondents interact and use the attributes employed within hypothetical Discrete Choice Experiments (DCE). For example, Hensher, Rose and Greene (2005) explain that it is normally assumed that when a survey participant undertakes a hypothetical DCE they employ all attributes. However, there are reasons to assume that respondents may employ less than the full set of attributes when making choices. Within the literature this form of behavior has become known as attribute non-attendance (ANA) and its existence has been shown to significantly effect model performance (Scarpa, Thiene and Hensher, 2010; Balcombe, Burton and Rigby, 2011; Scarpa et al., 2013). To date two general approaches have developed to examine ANA. Either debriefing questions are included between choice sets (Scarpa, Thiene and Hensher, 2010; Puckett and Hensher, 2009) or at the end of the choice sets (Campbell, Hutchinson and Scarpa, 2008). Inclusion at the end of choice sets has been more widely employed in practice. Debriefing questions directly ask respondents which attributes they used or did not. Alternatively, econometric methods have been employed to reveal ANA *ex-post* from a data set (Scarpa et al., 2009; Hensher, Rose and Greene, 2012). This approach is often referred to as a form of post estimation conditioning. Generally, most studies focus on one approach or the other, although Hess and Hensher (2010) do provide an interesting comparison of both approaches.

A central issue within the stated ANA literature has largely been on whether respondents really ignore attributes and what the implications of this would be for Random Utility Models. It is now well known that many respondents, when prompted, often state that they ignore some subset of the attributes presented to them in a hypothetical DCE. For example Campbell, Hutchinson and Scarpa (2008) report that 36 percent of respondents do not use at least one attribute. So while the initial goal of the ANA literature was to determine whether people have employed simplification strategies, this literature has resulted in demonstrating that asking debriefing questions about attribute attendance is an important source of information about peoples utility functions. However, with exceptions (Balcombe, Burton and Rigby, 2011), the majority of papers seem to suggest that respondents do not fully ignore attributes that they state that they do not attend. Essentially, it seems that respondents that indicate ANA place lower importance, which need not be zero, on those attributes when making choices, but they do not ignore them altogether (Hess and Hensher, 2010). If stated non-attendence is an indicator of an attribute's 'value', asking respondents if they have ignored an attribute with a simple dichotomous yes/no question might be views as a crude approach. A non-attendence response no longer signals a zero value on the contribution of a specific attribute within the econometric model, and setting the marginal utility to zero, as is typically done, may impair model performance.

In this paper we take a different approach to stated ANA. Instead of asking people whether they have ignored (or used) attributes within our hypothetical DCE we ask them to rank the attributes in order of importance to them. This should not be confused with a ranking approach for alternatives that is reasonably common within the DCE literature (Layton, 2000; Watanabe, 2010; Scarpa et al., 2011). As with much of the existing ANA literature our ranking question is employed after all the choices have been completed. We also note that there is no reason *a priori* that our approach could not be implemented after each choice set. By employing a single ranking question we allow survey respondents to place a lower value on particular attributes without assuming that they have zero value. Since respondents only perform this task once, this simple de-briefing question offers important insights into respondent behavior with only a small increase in the total cognitive burden placed on respondents. We show how this information can be used in a parsimonious way by modifying the Mixed Logit without imposing the condition that the ranking information must necessarily indicate an attributes relative marginal utility.

Overall, we believe that our approach provides an interesting alternative to the assessment of attribute use and importance compared to a dichotomous non-attendance question. As our results demonstrate the inclusion of attribute rank data within the model significantly improves model performance. However, we also acknowledge that by asking respondents to rank the importance of attributes that we do not in principle explicitly reveal attribute non-attendance. But, as previously noted within the literature (Hess and Hensher, 2010) simply offering a respondent a yes or no option ignores the possibility that a specific attribute only has lower importance as opposed to zero importance.

Here we examine two alternative ways of incorporating ranking data. The first uses the ranking data as a covariate. The second, which is new to the literature, uses the ranking data to scale the parameters in a manner we will refer to as the "contraction" approach. We first assess if rank data are consistent with marginal utilities estimated independently of the ranking data. We then employ a modified (Bayesian) Mixed Logit model that incorporates the ranking data and we make model comparisons employing model marginal likelihoods.

Our specific application is a hypothetical DCE study into the attributes of bread, including a functional ingredient and a health claim. The inclusion of both attributes was employed to allow us to examine the relative importance of each attribute for survey participants. As such this DCE adds to a growing literature examining consumer attitudes towards foods modified with functional ingredients as well as the provision of information to help consumers make informed food choices (Cowburn and Stockley, 2005; Grunert and Wills, 2007; Mazzocchi, Traill and Shogren, 2009; Balcombe, Fraser and di Falco, 2010; Hellyer and Haddock-Fraser, 2011; Hellyer, Fraser and Haddock-Fraser, 2012).

The hypothetical DCE employed in this paper has previously been analyzed by Bitzios, Fraser and Haddock-Fraser (2011). However, we extend the previous analysis by employing attribute ranking data as well as 318 additional survey responses collected online. As the DCE collected data using two modes of survey delivery – mail and online, we are able to compare model performance for both types of data. There already exist several studies that examine if the mode of DCE survey delivery impacts resulting model estimates (e.g., Savage and Waldman, 2008; Olsen, 2009; Lindhjem and Navrud, 2011; Windle and Rolfe, 2011).

Our analysis adds to this literature by examining if there are differences in model results for the mail and online survey data for all models estimated.

The structure of the paper is as follows. In section 2 we briefly describe the hypothetical DCE employed in this study. We then introduce and develop the econometric models we use to estimate our data. In section 4 we describe our data and report model results. In Section 5 we provide a summary and conclude.

DCE Design and Data

The hypothetical DCE employed in this paper was designed to provide willingness-to-pay (WTP) estimates for various types of bread with assorted attributes. The data employed had two modes of delivery, a mail version and an online version. Bitzios, Fraser and Haddock-Fraser (2011) analyzed the mail version data only using a latent class approach, and did not employ the ranking data as we do in this paper. The two versions of the survey employed in this paper only differ in their mode of delivery. A full description of the design of the DCE can be found in Bitzios, Fraser and Haddock-Fraser (2011) including the approach to attribute selection, experimental design and choice card format. A brief description of the attributes and levels employed in the DCE are provided in Table 1.

[**Approximate Position of Table 1**]

The survey had four different versions (24 options that were presented to respondents in four blocks of six choice cards). The survey was composed of six sections. The first section gave information and explained the concept of functional foods and contrasted these to a typical health claim with an associated benefit. Both concepts were defined in the survey instrument based upon agreed rules governing claims on food products in the UK. The second section included some warm-up questions on bread eating behavior and bread knowledge. The third section explained the choice task using an example, and the fourth section presented the actual choice exercises that had to be completed. The next section included questions about attitudes towards food. In addition, this section included the ranking of attributes question. The final section collected socio-economic individual specific information.

The specific ranking question that we asked was as follows:

For your choice card responses please rank from 1 (Most Important) to 7 (Least Important) the attributes which affected your choices. No two attributes should receive the same rank number.

- *Type of bread*
- *Production method of grain*
- *The presence of functional ingredient*
- *Whether it is sliced or unsliced*
- *The texture of bread*
- *The potential health benefit*
- *Price of bread*

The online version of the survey was implemented using SurveyMonkey an online survey software and questionnaire tool (www.surveymonkey.com/). We employed an opt-in approach to survey participation. To attract survey participants we placed a link to the survey on the University of Kent website, advertised via the news section of the University's website. The advertisement provided a link for respondents to the survey. We also placed a link on the Home Grown Cereals Authority website which was advertised via their e-club "Crop Research News". For both sites the link to the specific version of the survey was modified every few days to ensure that we obtained a balance of responses across the four blocks of choice cards we had employed with the postal version of the survey instrument. The mail survey had 341 usable responses and the online survey returned 318. A comparison of both mail and online respondents is provided in Table 2.

[**Approximate Position of Table 2**]

Table 2 shows that there are a number of statistical differences in the two samples. For example, we have more female respondents than males for both survey modes, and that the proportion of females is significantly higher for the online version of the survey. The actual proportion of females in the UK is just under 51 percent. Our mail sample has an above average age compared to the UK average of 39, whereas the online sample has a lower average age. The average income of respondents (excluding non-responses) is just over £31,000 for mail and £33,000 for online which is reasonably close to average income in the UK.

Notably, the online survey attracted proportionally more females than the mail survey and generally the online participants were considerably younger. The online participants also tended to be slightly more highly educated, paid and in work, and health conscious.

In terms of the attribute ranking raw data presented in Table 2 it is evident that type of bread is clearly identified as the most important attribute by respondents for both survey modes. Also we note that the sample average score for both groups is significantly different. This is followed by price, texture, and health benefit. Interestingly, the statistical significance of the mean score differences between the survey modes is less for these three attributes compared to those attributes that are ranked lower. As we might expect an explicit health claim in the form of a benefit ranks higher than the inclusion of functional ingredient which may yield health benefits. Despite some of the identified differences in sample composition the rank order of DCE attributes was the same across the two modes of delivery.

In Section 4 the importance rankings will be used within the estimation of the Mixed Logit. As we will see these rankings are able to be used in the estimation of marginal utilities and they do have an impact.

Model Specification and Estimation

The Standard 'Mixed Logit' (Model 1)

The utility (U) that the jth $(j = 1, ..., J)$ individual receives from the ith choice $(i = 1, ..., I)$ in the sth choice set $(s = 1, ..., S)$ is assumed to be of the form

$$U_{ijs} = \dot{x}'_{ijs}\dot{g}\left(\beta_j\right) + e_{ijs} \tag{1}$$

where $\dot{x}_{ijs}$ denotes the $K \times 1$ vector of attributes presented. The error e_{ijs} is 'extreme value' (Gumbel) distributed, is independent of $\dot{x}_{ijs}$, and is uncorrelated across individuals or across choices. β_j is a $(k \times 1)$ vector describing the preferences of the *jth* individual and obeys

$$\beta_j = \alpha + u_j \tag{2}$$

where α is the mean and u_j is a independently and identically normally distributed vector with variance covariance matrix Ω. The function $\dot{g}(\beta_j) = (\dot{g}_1(\beta_{1j}), ..., \dot{g}_K(\beta_{Kj}))$ is a dimension preserving transformation of the vector β_j. For example, by using a exponential transformation for a given attribute coefficient, the marginal utility for that attribute becomes log normal. The errors $\{u_j\}$ are assumed to be uncorrelated across individuals. It is also common to condition the marginal utility in (2) on variables that characterize the respondent, as we discuss below.

Ranking as Covariates (Model 2)

In this DCE we have observations $\{z_{jk}\}$ which represent the rank of the kth attribute by the jth respondent. As outlined above, each respondent was required to rank the data on a scale from one through R (in case $R = 7$). Respondent were required to assign a unique rank to each attribute (with no ties allowed) with one being the highest ranked (most important) attribute and R being the lowest. Note, in the case where a given attribute is categorical so that the coding uses dummy variables then the number of attributes to be ranked (R) will be smaller than K. Each of the dummy variables associated with a given attribute will receive the same rank.

In common with the treatment of non-attendance data, we could choose to extend (2) so as to treat the rank as an explanatory variable for β_j. More specifically

$$\beta_j = \alpha_0 + \alpha_1 z_{jk} + u_j \tag{3}$$

In equation (3) α_0 is equal to α in equation (2) if α_1 is equal to zero which occurs if the ranking data has no impact on the model. However, if the rank data does impact the model then α is equal to $\alpha_0 + \alpha_1$ times z_{jk} at its mean. Note, we only report α (the net effect) for this model as this is the value used to derive our estimates of WTP.

This 'covariate approach' is potentially unsatisfactory because by treating the variance term of β_j as invariant to the ranking of an attribute we ignore the fact that it is not only a shift in the mean that would be expected but that people with very low rankings of some attributes are more likely to have marginal utilities clustered around zero.

The Contraction Appproach (Model 3)

In order to take account of the problems identified with the use of the attribute ranking data in model 2 we now propose an alternative, where we define utility as in (1).

First, let us define the matrix $\Lambda_j = diag(\lambda_{j1}, ...\lambda_{jK})$ which has the elements

$$\lambda_{jk} = (1 - \tau) + \tau \frac{(R - z_{jk})}{R - 1} \tag{4}$$

where τ is a parameter that is to be estimated and is free to vary between zero and one. As $\tau \to 0$ this implies that the ranking data is unimportant in determining the mean and variance of the coefficients. At the other extreme, $\tau = 1$ implies that the lowest ranked attribute has zero marginal utility. How does this work? If we assume that $\tau = 1$ and $R = 7$ and $z_{jk} = 7$, then by substituting these values into (4) that yields a value of $\lambda_{jk} = 0$. In this case this implies that α is equal to zero for the lowest ranked attribute. In contrast, if assume that $\tau = 0.5$, $R = 7$ and $z_{jk} = 6$, and again substituting these values into (4) we now find that $\lambda_{jk} = 0.583$ which implies that the ranking data is important and that it influences yields an estimate of α equal to based on $0.583 * \alpha_0$ where α_0 is equal to α in equation (2). Clearly, the higher the (mean) rank of an attribute the bigger the relative estimate of λ_{jk} and the lower the contraction effect on the resulting estimate of α. Finally, as with the covariate model we only report the estimates for final estimates of α as they are used to calculate WTP.

It then follows that the individual marginal utilities are modelled by assuming $g\left(\beta_j\right) = \left(g_1\left(\beta_{j1}\right), \ldots.., g_K\left(\beta_{jK}\right)\right)$ where g_k is a transformation (e.g. an exponential) and likewise defining the elements of $\dot{g}\left(\beta_j\right)$

$$\dot{g}_k\left(\beta_{jk}\right) = \lambda_{jk} g_k\left(\beta_{jk}\right) \tag{5}$$

We note that for the highest ranked attribute $\lambda_{jk} = 1$ regardless of the value of τ. Without this condition the model would not be identified. We note that a similar condition is employed by Layton (2000) in his examination of DCE rank data. We refer to this model format as the 'contraction approach'. We can write this in vector form using

$$\dot{g}\left(\beta_j\right) = \Lambda_j g\left(\beta_j\right) \tag{6}$$

Estimation of the Contraction Model

The contraction model is simple to estimate using Bayesian methods, since it can be specified in a similar way to the standard Mixed Logit, with the normal latent variables being multiplied by the shrinkage terms. If we define

$$g\left(\beta_j\right) = \Lambda_j^{-1} \dot{g}\left(\beta_j\right) \tag{7}$$

where as before:

$$\beta_j \sim N\left(\alpha, \Omega\right) \tag{8}$$

Viewing utility in this way we have

$$U_{ijs} = \left(\dot{x}'_{ijs}\Lambda_j\right) g\left(\beta_j\right) + e_{ijs} \tag{9}$$

By defining

$$x'_{ijs} = \dot{x}'_{ijs}\Lambda_j \tag{10}$$

the non-stochastic component of utility is defined conventionally as

$$V_{ijs} = x'_{ijs} g\left(\beta_j\right) \tag{11}$$

and the posterior densities for the parameters $\left\{\beta_j\right\}, \alpha, \Omega$, and τ, are obtained by observing that the probability of i being chosen in the circumstance js is the standard logit probability

$$p_{ijs} = \frac{e^{V_{ijs}}}{\left(\sum_i e^{V_{ijs}}\right)} \tag{12}$$

If the observed choices are defined by $y_{ijs} = 1$ where the ith option is chosen in circumstance js and $y_{ijs} = 0$ otherwise, then the likelihood of all the observed choices (Y) is

$$f\left(Y|\tau, \alpha, \Omega\right) = \prod_i \prod_j \prod_s p_{ijs}^{y_{ijs}} \tag{13}$$

Conditionally on Λ_j, the steps for generating latent variables $\left\{\beta_j\right\}$ along with α and Ω can be estimated using Markov Chain Monte Carlo (MCMC) steps as in the standard Mixed Logit (e.g. Train and Sonnier, 2005). That is, having normalized the attributes $\left(x'_{ijs} = \dot{x}'_{ijs}\Lambda_j\right)$ the conditional distributions for β_j along with α and Ω are defined in the usual way (in terms of x_{ijs}). However, since τ is estimated, the normalized attributes need to be updated at each iteration, and the posterior distributions for τ is also required. The precision matrix has a Wishart prior $W\left(I, k+4\right)$ where k is the dimension of the covariance matrix. The precise priors that we use have a mean of zero for α and a diagonal covariance matrix for α with a variance of 100 for each of the effects common to all models. For the covariate terms in the model using the ranking data (model 2) the variances were set to 10. Thus, the prior variance for α was set so as to be relatively uninformative for the estimates, and small enough so that the penalty for additional parameters in the model would not be very restrictive. Therefore, it follows that the posterior distributions for τ is

$$f(\tau|Y,\alpha,\Omega) \propto f(Y|\tau,\alpha,\Omega) f(\tau) \tag{14}$$

where τ has a uniform prior over the unit interval [0,1]. Estimation proceeds by iterating through the sequence of conditional draws: $\{\beta_j\} |\alpha,\Omega,\tau,Y; \alpha| \{\beta_j\},\Omega,\tau,Y; \Omega| \{\beta_j\},\alpha,\tau,Y;$ $\tau|\alpha,\Omega,\{\beta_j\},Y$. The conditional posterior distributions for the first three components are the same as in Train and Sonnier (2005). The conditional posterior distribution for τ is obtained from (14). These can be sampled using Metropolis Hastings steps with a random walk proposal density.

Results

Model Comparisons

We now examine the relative performance and results of three competing models across the two data sets (Mail and Online). The three models which we employ differ in their treatment of the ranking data. The first model (Model 1) makes no use of the ranking data. The second model (Model 2) uses the ranking data as a covariate on marginal utilities, thus allowing the mean to depend on the rankings of attributes (as in (3)). The third model (Model 3) uses the ranking data in the manner described previously.

The results for the logged marginal likelihoods (MargLL) are presented in Table 3.

[Approximate Position of Table 3]

For completeness we also present the maximum log likelihood (MaxLL) (calculated using the simulation method with Halton Sequences) visited by the sampler. From a Bayesian perspective the MargLLs are sufficient for us to make model comparisons (Balcombe, Fraser and Chalak, 2009). Comparisons should only be made vertically (we are not comparing between online and mail surveys). The larger the MargLL, the 'more preferred' a model. The exponential of the difference between the MargLL for two models gives the 'Bayes

Factor' between two models when each is considered equally plausible *a priori*. For example, models which have a difference of three in the MargLL would indicate that the model with the larger MargLL is over 20 times more likely to be the true model after incorporating the sample information. The MargLL implicitly takes into account whether one model has more parameters than another, so no adjustment needs to be made to the MargLL in order to make model comparisons.

As the results show, in most cases the differences between the MargLLs between competing models are quite large. For both the mail and online data Model 3 is preferred to Model 2 which in turn is preferred to Model 1. As can also be seen from the MaxLL within Table 3, there is also a very large improvement in the MaxLL when comparing Model 3 with Model 1, even though there is only one additional parameter. Since Model 3 nests Model 1, one could calculate a classical p-value using a likelihood ratio statistic that would reject the restriction that $\tau = 0$ at very low levels (p<0.001). The results, therefore, seem unequivocal. Using the ranking data improves model performance whether ranks are used as covariates, or the contraction approach. However, as can also be seen there is a large improvement in MargLLs from using the contraction approach over the covariate approach.

Standard Mixed Logit (Model 1)

We first present the results of the parameter estimates of the standard Mixed Logit (Model 1) in Table 4.

[**Approximate Position of Table 4**]

We consider this model because our first interest is about whether there is a relationship between the importance rankings (reported in Table 1) and the size of the coefficients when they are estimated independently of the ranking data. Within Table 4 we report, for both online and mail data, the estimates and standard deviation of α (in columns 1, 2, 4 and 5) along with the estimates (the mean of the posterior) for the diagonal elements of Ω (in columns 3 and 6). These are referred to as 'the mean of the variances'. Whereas α

determines the means of the latent variables, the variances Ω_{ii} determine how diffuse these marginal utilities are across the population. If $\sqrt{\Omega_{ii}}$ is large relative to α_i (unless the utility is transformed) then a significant part of the population will have differently signed marginal utilities.

As can be seen from Table 4 the average attribute importance scores reported in Table 2 correspond reasonably with the size of the coefficients which, given that they are mainly dummies, are able to be compared. This is most evident with regard to the bread type. We see that whether a bread is wholegrain or brown has a very large average marginal utility, though this does differ substantially across the population (the variance estimates reflecting respondent heterogeneity are high). Examining the importance rankings in Table 2 we see that bread type was considered the most important attribute on average. Likewise, the next most important attribute (texture) also seems to have a relatively large effect on peoples utility given the coefficients in Table 4. The fourth most important ranked attribute is the health benefit which seems to play a large role in peoples choices given the quite large marginal utility (0.819) and relatively small standard deviation for this estimate (0.112). Importantly, for both survey modes health claims yield higher levels of marginal utility compared to a functional ingredient. This in part goes back to the difference in these attributes. As defined in the survey instrument; *"Functional foods are products that, as part of a healthy diet, promote health and help reduce the risk of certain diseases."* In contrast food with a health claim was defined as; *"Healthy foods are beneficial for the general state of your health."* Thus, with a functional ingredient there is a conditional relationship between consumption of the food and a positive health outcome. In contrast a health claim makes an explicit and general link between consumption and health.

Finally, if we compare the results across survey mode we see that there are few significant differences in sign, although these tend to be associated with α_i estimates that have a relatively high standard deviation e.g., method of production and thick sliced. We note the high mean of the variance for rye bread which indicates that respondents typically either really like or dislike this type of bread.

Rankings as Covariates (Model 2)

We now examine the impact of the attribute ranking data when they are included as covariates on the marginal utilities. These results are presented in Table 5.

[Approximate Position of Table 5]

From Table 5 we can see that the importance rankings seem to be strongly correlated with the marginal utilities. For this exercise we took the (1 to 7) score as the covariate so as to make the results easier to interpret. We would expect that marginal utility which was positive would have a 'significant' positive ranking coefficient. As we can see for bread types, price and health benefit, this is indeed the case.

There are a couple of counter intuitive results. First, is texture, whereby although the effects included in the models were positive, those indicating that they have high importance for these attributes were estimated to have lower utilities (as shown by the fact that the dummy covariates have negative signs). This result is consistent across both survey modes. It is likely that this result highlights the fact that the type of texture coded as the base level (i.e., soft) is the generally preferred type of this attribute. Second, the method of production is now positive for the mail survey model and relatively more important than functional ingredients. Third, there is a reversal in signs for the sliced attribute estimates. However, the magnitude of the associated standard deviations for the α estimates indicates that these estimates need to be treated with caution.

Contraction Model (Model 3)

We now present our estimates of Model 3 using the contraction approach. These results are shown in Table 6.

[Approximate Position of Table 6]

The first thing to note are the contraction coefficient estimates at the bottom of Table 6. The estimates for the contraction coefficients are approximately 0.94 and 0.80 for the mail and online versions respectively. These estimates are high suggesting that people have very small marginal utilities for those attributes they rank as having low importance. Also, for both survey modes these estimate are statistically significant.

In terms of interpretation, the coefficient of 0.94 for the mail version of the survey indicates that for a respondent who ranks a specific attribute the lowest (i.e., 7th), then they would have a marginal utility of 6% (0.06 derived from equation (4)) of that which they would otherwise be predicted to have. For the online version the lowest ranked attribute would have a marginal utility of 20%.

It we consider higher ranked attributes, a higher rank score will mean that the impact of the contraction coefficient is reduced. So for an attribute ranked third most important, using the estimates reported in Table 6 and equation(4), for the mail version the associated marginal utility will be 69%, whereas for the online version the marginal utility will be 73%.

Overall, while both surveys give comparable results, those in the mail version have a significantly greater contraction coefficient. This in part might be as a result of the greater spread of mean ranks scores that are reported in Table 2. As we can see in Table 2 the mail survey yields the highest and lowest average rank scores recorded.

Turning to the estimates of marginal utility there is a reasonable correspondence between mail and online for most attributes, except for differences between rye, crunchy and springy. As with the covariate model texture yields some negative estimates for the mail version, although these are all positive for the online version. As above it is likely that the type of texture coded as the base level (i.e., soft) is the preferred type.

WTP Estimates

We need to be clear that the values of α within Tables 4, 5 and 6 cannot be directly compared. It is possible to obtain a rescaling of the α coefficients at the mean. However, this can effectively be done through the WTP estimates. As such we now present in Table

7 the WTPs for all three models.

[Approximate Position of Table 7]

The WTPs are estimated using simulation from the distribution of the latent coefficients and contraction coefficients. In Table 7 we see that the estimates are, for the most part, fairly robust to changes in method and survey mode.

If we compare Models 1 and 3, we can see that there is a tendency for downward absolute revision in WTP estimates, although the changes are not dramatic. For example, for the wholegrain estimates the reduction is 14 percent for the mail survey and 10 percent for the online survey. However, this was not the case where the attribute rank score was used as a covariate (Model 2). In this case the WTP estimates tended to become slightly higher.

According to these results, it is striking that people are prepared, on average, to pay a large premium for wholegrain breads (anywhere from around £1.46 to £2.18) taking the lowest and highest estimates. However, the best performing model (Model 3) gives the lowest estimates (£1.46 to £1.76 mail or online respectively).

The most noticeable difference between the mail and online results is in the WTP results for method of production: conventional versus organic. For the mail results we found very small or even negative WTP for organic bread, whereas this result was given a premium of 30 pence for the online. Slightly larger values were also found online for the inclusion of a functional ingredient and for a health benefit. Over all models and survey modes, the health benefit was given a higher WTP than for the functional ingredient or organic production, with an estimate of an average 60 pence premium for the health benefit. We also note that the respondents appear more homogeneous in their liking for the health benefit, whereas there was a great deal of heterogeneity across the population about liking for organic production or functional ingredients.

Summary and Conclusions

This paper has introduced a new way of using respondent debriefing ranking information about attribute importance in the context of a hypothetical DCE for various attributes of bread. The ranking information was incorporated into the Mixed Logit using a new model specification. Our results indicate that a DCE debriefing question that asks respondents to rank the importance of attributes helped to explain the resulting choices and improved estimates of respondent utility functions. We explored incorporating the ranking information in two different ways: as a covariate explaining marginal utilities and a 'contraction' of the marginal utility towards zero where the degree of contraction was estimated. The second approach proved to be the preferred one in terms of overall model performance, although the covariate approach also improved model performance relative to using no information at all. The mode of survey delivery (online and mail) did not substantively alter our conclusions either with regard to the use of debriefing information or with regard to the estimates of marginal utilities and WTP. The results indicated that attributes which were ranked the lowest by respondents had a very small marginal utility for those respondents.

With regard to the determinants of people's WTP for attributes of bread, the largest premiums were, on average, attached to 'wholegrain' closely followed by 'brown', but with a very large variation across the population with many consumers preferring white bread. Organic production received only a small premium on average, as did 'functional ingredients'. However, a health benefit in the form of claim was valued highly by the vast majority of the survey respondents.

The research in this paper has built upon the literature on stated ANA which has shown that debriefing questions about attribute knowledge can assist our understanding of respondent utility functions in a way that is complementary to the observation of discrete choices. Overall the ranking exercise undertaken by respondents is a relatively low cost exercise and we would advocate its use in DCE.

More generally there is good reason to assume that the results regarding contractions based on rankings may depend, *inter alia*, on the number of attributes in the DCE. There is already an interesting literature developing on the complexity of DCE and in particular the number of attributes (Burton and Rigby, 2012). We believe that combining work on design complexity along with the type of debriefing questions examined in this paper and the econometric methods is an area of research that warrants further investigation. There is also further work to be done on how best to formally incorporate other forms of information into the estimation process using multiple debriefing questions. For example, as Scarpa et al. (2013) note, it would be interesting to see if respondent eye-tracking data collected during the choice process could be used to explain attribute use. Preliminary results, reported in Balcombe, Fraser and McSorely (2013) appear to support this conjecture about the potential of using eye-tracking to enhance data collection and subsequent model performance for DCE.

Finally, the question of endogeneity within discrete choice models has recently been raised by a number authors (e.g. Petrin and Train, 2010, Hess and Hensher, 2012) and this is potentially of relevance to the type of problem being examined in this paper. We have two comments about the implications of this issue for this paper.

First, the type of endogeneity addressed by Petrin andTrain (2010), is the question of whether the explanatory variables in an equation are independent of the error term. For example, if price is correlated with missing variables (such as advertising) then if it is used as an explanatory variable in a utility or demand equation, then endogeneity bias will result. Petrin and Train (2010) introduce the control variable approach to deal with this problem. While clearly important in context of non-experimental revealed preference studies where one cannot control for non-observed effects, the applicability of this approach is less relevant to a hypothetical stated preference setting where researchers try and frame the choices so as to explicitly preclude endogeneity. It may be argued that respondents in a DCE continue to infer the existence of attributes not included in the DCE, on the basis of those that are included, and as such infer the actual level of these attributes using the levels of those in the experiment. However, in such a setting it would be difficult to design instruments that could remove the endogeneity bias. Rather, the correct route would seem to be to get respondents to understand that they should assume that all other attributes that are not explicitly included in the experiment should be treated as equal across the options.

Second, it is possible that the source of endogeneity pertains to the possibility of factors that determine both the random errors in the utility equation and the preference parameters for individuals (it does not posit endogeneity between the attribute levels and the utility errors as in the case of Petrin and Train (2010)). With regard to this problem, Hess and Hensher (2013) have suggested that there may be potential endogeneity problems related to the use of supplementary questions, such as ANA questions or attribute ranking questions employed in this paper. However, it should be recognised that within the context of models such as the Mixed Logit, there is an assumption that the individual specific preference parameters are independent of the Gumbel errors in the utility equation. Should this fail to be the case then there is endogeneity, but in this case endogeneity bias will be present with or without the use of supplementary questions in estimation. While the approach in Hess and Hensher (2013) is a potentially useful way of approaching the use of non-attendance and/or rankings data, we do not believe that it really relates to the issue of endogeneity. Hess and Hensher (2013) make the same assumptions as in the standard Mixed Logit, that the rankings data is independent of the Gumbel error, and that it drives the individual preference parameters. In that respect it is not very different from what we have done in this paper.

Acknowledgements

We acknowledge the financial support provided as part of a collaborative project (SLOW-CARB) involving the Home Grown Cereals Authority, Orafti, Premier Foods, Naturis, the University of Surrey and Imperial College London. We also thank Brian Roe and two referees for comments on an earlier version of this paper.

References

Balcombe, K.G., Burton, M. and Rigby, D. (2011). Skew and Attribute Non-Attendance within the Bayesian Mixed Logit Model, *Journal of Environmental Economics and Management*, 62(3): 446-461.

Balcombe K. Fraser I.M. and Chalak A. (2009). Model Selection in the Bayesian Mixed Logit, *Journal of Environmental Economics and Management*, 57(2): 226-237.

Balcombe, K., Fraser, I.M. and di Falco, S. (2010). Traffic Lights and Food Choice: A Choice Experiment Examining the Relationship Between Nutritional Food Labels and Price, *Food Policy*, 35(3): 211-220.

Balcombe, K., Fraser, I.M. and McSorely, E. (2013). Visual Attention and Attribute Attendance in Multi-Attribute Choice Experiments, Memo.

Bitzios, M. Fraser, I.M. and Haddock-Fraser, J. (2011). Functional Ingredients and Food Choice: Results from a Mixed Mode Study Employing Means-End-Chain Analysis and a Choice Experiment, *Food Policy*, 36(5): 714-724.

Burton, M. and Rigby, D. (2012). The Self Selection of Complexity in Choice Experiments, *American Journal of Agricultural Economics*, 94(3): 786-800.

Campbell, D., Hutchinson, W.G. and Scarpa, R. (2008). Incorporating Dis-continuous Preferences into the Analysis of Discrete Choice Experiments, *Environmental and Resource Economics*, 41: 401-417.

Cowburn, G. and Stockley, L. (2005). Consumer Understanding and Use of Nutrition Labelling: A Systematic Review, *Public Health Nutrition*, 8(1): 21-28.

Grunert, K.G. and Wills, J.M. (2007). A Review of European Research on Consumer Response to Nutrition Information on Food Labels, *Journal of Public Health*, 15: 385-399.

Hellyer, N., Fraser, I.M. and Haddock-Fraser, J. (2012). Food Choice, Health Information and Functional Ingredients: An Experimental Auction Employing Bread, *Food Policy*, 37(3): 232-245.

Hellyer, N. and Haddock-Fraser, J. (2011) Reporting diet-related health issues through newspapers: portrayal of cardiovascular disease and Type 2 diabetes. *Health Education Research*, 26 (1): 13-25

Hensher, D.A., Rose, J. and Greene, W.H. (2005). The Implications on Willingness to Pay of Respondents Ignoring Specific Attributes, *Transportation*,32: 203-222.

Hensher, D.A., Rose, J.M. and Greene, W.H. (2012). Inferring Attribute Non-Attendance from Stated Choice Data: Implications for Willingness to Pay Estimates and a Warning for Stated Choice Experiment Design, *Transportation*, 39: 235-245.

Hess, S. and Hensher, D.A. (2010). Using Conditioning on Observed Choices to Retrieve Individual-Specific Attribute Processing Strategies, *Transportation Research B*, 44(6): 781-790.

Hess, S. and Hensher, D.A. (2013). Making Use of Respondent Reported Processing Information to Understand Attribute Importance: A Latent Variable Scaling Approach, *Transportation*, (forthcoming).

Layton, D.F. (2000). Random Coefficient Models for Stated Preference Surveys, *Journal of Environmental Economics and Management*, 40: 21-36.

Lindhjem, H. and Navrud, S. (2011). Using Internet in Stated Preference Surveys: A Review and Comparison of Survey Modes, *International Review of Environmental and Resource Economics*, 5: 309-351.

Mazzocchi, M., Traill, B,W. and Shogren, J.F. (2009). *Fat Economics. Nutrition, Health, and Economic Policy*, Oxford University Press.

Olsen, S.B. (2009). Choosing Between Internet and Mail Survey Modes for Choice Experiment Surveys Considering Non-Market Goods, *Environmental and Resource Economics*, 44: 591-610.

Petrin, A. and Train, K. (2010). A Control Function Approach to Endogeneity in Consumer Choice Models, *Journal of Marketing Research*, XLVII: 1-10.

Puckett, S.M. and Hensher, D.A. (2009). Revealing the Extent of Process Heterogeneity in Choice Analysis: An Empirical Assessment, *Transportation Research A*, 43(1): 117-126.

Savage, S.J. and Waldman, D.M. (2008). Learning and Fatigue During Choice Experiments: A Comparison of Online and Mail Survey Modes, *Journal of Applied Econometrics*, 23: 351-371.

Scarpa, R., Gilbride, T.J., Campbell, D. and Hensher, D.A. (2009). Modelling Attribute Non-Attendance in Choice Experiments for Rural Landscape Valuation, *European Review of Agricultural Economics,* 36(2): 151-174.

Scarpa, R., Thiene, M. and Hensher, D.A. (2010). Monitoring Choice Task Attribute Attendance in Nonmarket Valuation of Multiple Park Management Services: Does it Matter? *Land Economics*, 86(4): 817-839.

Scarpa, R., Notaro, S., Louviere, J. and Raffaelli, R. (2011). Exploring Scale Effects of Best/Worst Rank Ordered Choice Data to Estimate Benefits of Tourism in Alpine Grazing Commons, *American Journal of Agricultural Economics*, 93(3): 813-828.

Scarpa, R., Raffaele, Z., Bruschi, V. and Naspetti, S. (2013). Inferred and Stated Attribute Non-Attendance in Food Choice Experiments, *American Journal of Agricultural Economics*, (Forthcoming).

Train, K. and G. Sonnier (2005). Mixed Logit with Bounded Distributions of Partworths, in *Applications of Simulation Methods in Environmental Resource Economics*, R. Scarpa and

A. Alberini (Eds.), Kluwer Academic Publishing.

Windle, J. and Rolfe, J. (2011). Comparing Responses from Internet and Paper-Based Collection Methods in more Complex Stated Preference Environmental Valuation Surveys, *Economic Analysis and Policy*, 41(1): 83-97.

Table 1: Attributes and Levels Employed in the DCE

Attributes	Description	Levels
Type of Bread	Breads offered in the hypothetical market	White, Wholemeal Brown, 50-50, Rye
Method of Production	Grain type used in bread	Conventional, Organic
Functional Ingredient	Ingredient that can potentially deliver nutritional benefits	Yes, No
Sliced/Unsliced	Bread sold sliced or not	Medium, Thick, Unsliced
Texture	Consistency of the bread	Soft, Firm, Crunchy, Springy
Health benefit	If bread promotes health	Yes, No
Price	Cost (£) of standard 800gr loaf	0.7, 1, 1.3, 1.6, 1.9, 2.2

Table 2: DCE Descriptive Statistics

Socio-Economics (Avg)	Units	Sample	Mail	Online	Difference
Gender	Female=1	0.71	0.64	0.81	-0.18***
Age	Years	44.27	52.66	33.65	19***
Children	Number	0.45	0.47	0.42	0.05
Education	1 to 5	2.27	1.72	2.9	-1.18***
Income	£000's	32.12	31.02	33.61	-2.59
Exercise Regularly	Yes = 1	0.6	0.62	0.58	0.04
Health Conscious	Yes = 1	0.72	0.69	0.76	-0.07**
Gluten Intolerance	Yes = 1	0.04	0.05	0.04	0.02
Work	Yes = 1	0.57	0.54	0.6	-0.06*
Rank Scores (1 high, 7 low)					
Bread Type		2.03	1.89	2.19	-0.3**
Production Method		4.99	5.2	4.76	0.44***
Functional Ingredient		5.13	5.29	4.96	0.33***
Sliced		4.24	4.11	4.37	-0.26**
Bread Texture		3.73	3.67	3.81	-0.14
Health Benefits		4.13	3.99	4.22	-0.23*
Bread Price		3.78	3.85	3.7	0.15

Note: Statistically significantly different at 1% (***), 5% (**) and 10% (*).

Table 3: Marginal Log Likelihoods and Max Log Likelihoods

	Mail		**Online**		
	MargLL	**MaxLL**	**MargLL**	**MaxLL**	**No. of Parameters**
Model 1	-2083.66	-1968.86	-2058.92	-1954.64	104
Model 2	-2061.48	-1901.44	-2057.56	-1904.61	117
Model 3	-1994.19	-1889.72	-2016.16	-1911.78	105

Table 4: Standard Mixed Logit Results (Model 1)

	Mail			Online		
	Mean α	**St Dev α**	**Mean Var**	**Mean α**	**St Dev α**	**Mean Var**
Price (log-normal)	-0.441	0.223	1.723	-0.287	0.258	2.552
Bread (White)*						
Wholegrain	2.328	0.310	13.304	1.761	0.238	6.104
Brown	1.506	0.252	7.911	1.297	0.221	5.712
50/50	1.232	0.207	2.224	0.931	0.210	1.431
Rye	-0.429	0.339	14.736	-0.039	0.288	12.208
Method Production	-0.090	0.111	0.395	0.385	0.105	0.500
Functional Ingredient	0.250	0.112	0.226	0.227	0.103	0.196
Sliced (Thin)*						
Thick	0.083	0.124	0.482	-0.050	0.114	0.281
Unsliced	-0.217	0.134	0.583	-0.244	0.126	0.509
Texture (Soft)*						
Firm	0.327	0.158	0.997	0.331	0.141	0.617
Crunchy	0.134	0.148	1.065	0.220	0.133	0.534
Springy	0.251	0.150	0.590	0.353	0.139	0.677
Health Benefits	0.819	0.112	0.437	0.579	0.107	0.430

Note: * - Attribute level in brackets are the base level for dummy coding

Table 5: Impact of Rank on Mixed Logit (Model 2)

	Mail		Online	
	Mean α	**St Dev α**	**Mean α**	**St Dev α**
Price (log-normal)	-0.773	0.124	-0.612	0.136
Bread (White)*				
Wholegrain	0.387	0.172	0.317	0.107
Brown	0.308	0.140	0.223	0.100
50/50	0.163	0.116	0.085	0.079
Rye	0.268	0.181	0.315	0.130
Method Production	0.204	0.059	0.185	0.046
Functional Ingredient	0.172	0.066	0.076	0.056
Sliced (Thin)*				
Thick	0.007	0.055	-0.043	0.048
Unsliced	0.010	0.061	-0.110	0.053
Texture (Soft)*				
Firm	-0.104	0.073	-0.145	0.068
Crunchy	-0.021	0.077	-0.059	0.067
Springy	-0.162	0.073	-0.180	0.069
Health Benefits	0.336	0.054	0.206	0.049

Note: * - Attribute level in brackets are the base level for dummy coding

Table 6: Model Results With Contraction (Model 3)

	Mail			Online		
	Mean α	**St Dev α**	**Mean Var**	**Mean α**	**St Dev α**	**Mean Var**
Price (log-normal)	-2.307	0.271	2.919	-0.779	0.259	2.680
Bread (White)*						
Wholegrain	2.838	0.324	16.598	2.215	0.279	7.849
Brown	1.830	0.274	10.831	1.579	0.264	7.649
50/50	1.563	0.228	3.131	1.186	0.244	2.504
Rye	-0.597	0.403	21.630	0.079	0.351	16.527
Method Production	0.284	0.262	1.858	0.807	0.194	1.220
Functional Ingredient	0.878	0.223	0.834	0.566	0.194	0.542
Sliced (Thin)*						
Thick	-0.083	0.195	1.184	-0.046	0.163	0.479
Unsliced	-0.439	0.207	2.053	-0.508	0.200	1.722
Texture (Soft)*						
Firm	0.155	0.232	1.868	0.370	0.203	1.222
Crunchy	-0.315	0.253	3.931	0.207	0.204	1.470
Springy	-0.069	0.213	1.293	0.287	0.207	1.335
Health Benefits	1.624	0.172	0.653	1.150	0.160	0.603
	Mean	**St Dev**		**Mean**	**St Dev**	
Contract Coefficient	0.938	0.037		0.794	0.064	

Note: * - Attribute level in brackets are the base level for dummy coding

Table 7: Median WTP Estimates

	Mail			Online		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Price (log-normal)	1.000	1.000	1.000	1.000	1.000	1.000
Bread (White)*						
Wholegrain	1.71	1.848	1.465	1.970	2.181	1.768
Brown	1.086	1.159	0.906	1.432	1.531	1.200
50/50	0.906	0.947	0.852	1.066	1.097	0.950
Rye	-0.296	-0.371	-0.237	-0.065	-0.079	0.058
Method Production	-0.061	-0.056	0.025	0.396	0.391	0.293
Functional Ingredient	0.172	0.212	0.112	0.212	0.251	0.193
Sliced (Thin)*						
Thick	0.056	0.004	-0.010	-0.051	-0.082	-0.017
Unsliced	-0.146	-0.217	-0.063	-0.226	-0.292	-0.197
Texture (Soft)*						
Firm	0.232	0.227	0.032	0.325	0.397	0.183
Crunchy	0.098	0.099	-0.025	0.213	0.219	0.103
Springy	0.181	0.197	-0.010	0.369	0.410	0.144
Health Benefits	0.596	0.633	0.502	0.554	0.648	0.595

Note: * - Attribute level in brackets are the base level for dummy coding