Several methods for dealing with scale confound and efficiency in stated preference data with an empirical illustration

Emily Lancsar[1*] and Jordan Louviere[2]

1. Business School (Economics) and Centre for Health Services Research, University of Newcastle upon Tyne.
2. Centre for the Study of Choice, School of Marketing, University of Technology, Sydney.
* Author for correspondence: Emily.Lancsar@ncl.ac.uk

Work in progress. Please do not quote without the authors' permission.

Abstract

This paper presents a number of methods designed to address two key issues in the estimation and interpretation of choice experiments: 1) statistical efficiency and 2) confounds between 'weight' and scale. The first method we term "Best Worst Choices" (BWC). BWC is a way to obtain extra preference information by asking respondents to choose their most and least preferred options in each choice set, and from successively smaller subsets of the choice set, until an implied ranking of the options is obtained. The ranking effectively increases the amount of usable choice data, thereby increasing the statistical efficiency of the estimates. A second group of methods address an important issue which, until recently, was overlooked in health economics. That is, the attribute (level) parameters estimated from discrete choice experiment response data are confounded with the underlying subjective scale of the utilities, and strictly speaking cannot be interpreted as the relative 'weight' or 'importance' of the attributes, as is frequently done in the health economics literature. The relative importance of each attribute requires commensurable measurement units; that is, one needs a common, comparable, scale. We discuss and compare three methods that allow one to make such comparisons: 1) partial log likelihood analysis; 2) calculation of the marginal rate of substitution; and 3) estimation of Hicksian welfare measures. An empirical application is used to demonstrate each method and compare them to the relative importance of attributes elicited from direct ranking and rating questions. The empirical application also illustrates use of BWC, which is compared with the results of a standard, single-choice per choice set approach. Implications for task design, analysis and interpretation of choice experiments are discussed.

Words: 5304

Acknowledgements

The data used to demonstrate the methods proposed in this paper come from a choice experiment designed to investigate preferences for cardiac arrest in which both Cam Donaldson and Gillian Currie are involved. The experimental design used in this study was developed by Leonie Burgess and Deborah Street. The authors benefited from discussions with Tony Marley on expanding Best-Worst choices. We also gratefully acknowledge the support of the Australian Research Council, Grant Number DP0343632, entitled “Modelling the Choices of Individuals.” Emily Lancsar is funded by the Health Foundation and an Overseas Research Scholarship.

1. Introduction

This paper presents a number of methods designed to address two key issues in the estimation and interpretation of choice experiments. The first is the confound between attribute 'weight' and scale and the second is enhancement of statistical efficiency by collecting additional choice information in each choice set.

A common objective of discrete choice experiments (DCEs) in health economics is to compare the relative importance, or 'weight' of attributes of the product or program under investigation. Most studies have done this by comparing the size and significance of the estimated parameters for the attributes of interest, also referred to as 'part worth utilities'. Unfortunately, however, such comparisons ignore the fact that the parameters are not directly comparable because the attribute (level) parameters estimated from DCE response data are confounded with the underlying subjective scale of the utilities. Thus, the utility estimates for levels of each attribute cannot be interpreted as the relative 'weight' or importance of that attribute.

In particular, although the estimated utility of each attribute level is measured on an interval scale, the origins and units of each scale differ across attributes. Apart from obvious differences in underlying physical attribute units, such as price measured in dollars, time in minutes/hours etc, qualitative attributes have no physical referents. For example, the attribute levels for 'provider of care' might be nurse, doctor etc. A key issue is that the distance between levels on different attributes need not have the same meaning, or equivalently, the scale location or difference in utility between the levels of different attributes generally will not have the same utility scale units. It is possible to equate the origins of each scale, but it is not possible to equate scale units; hence direct comparisons of ranges of estimated utility parameters are not meaningful without transforming the data in some theoretically acceptable way. Put simply, one cannot determine whether the magnitude of the parameter estimate for an attribute level, and hence the resulting range of parameter levels for an attribute, is due to the 'importance' of this attribute or due to the position of each of the attribute levels on the underlying latent utility scale.

This can be seen by noting that the utility function of any choice model can be specified in latent utility space or in attribute space. For example, consider a case in which there are six attributes and all have L levels, with three attributes numerical and three attributes qualitative/categorical. One can specify the utilities equivalently by using effects or dummy codes to estimate the implied 6 x (L-1) utility values for L-1 levels of each attribute. Equivalently, one can specify the utilities by allowing for linear, quadratic, cubic, etc terms for each numerical attribute and effects or dummy codes for the qualitative attributes. If one uses L-1 polynomial codes for each numerical attribute, the degrees of freedom of the two specifications will be the same, and the statistical results also will be identical (except for rounding errors). We refer to the specification based entirely on effects or dummy codes as a specification in latent utility space, and the specification in polynomial codes (including any reduced form, such as specifying only linear numerical attribute effects) as a specification in attribute space. Hopefully, it is obvious that the two spaces are transforms of one another, allowing use of either depending on what one wishes to learn from the data.

Consider the estimated utilities of the L-1 qualitative attributes, leaving aside the numerical attributes for the moment. We can write this without loss of generality as

$$U_{ijk}\ldots = k_0 + k_1 u_i + k_2 u_j + k_3 u_k + \ldots, \qquad (1)$$

where $U_{ijk}\ldots$ is the latent overall utility for the attribute level combination that consists of the i-th level of attribute 1, the j-th level of attribute 2, the k-th level of attribute 3 and so forth; u_i, u_j and u_k are the respective utility values of the associated levels of attributes 1, 2 and 3, and k's are scaling constants. Let us now consider the properties of this expression by averaging the expression over the levels of attribute 1 (i.e., average over levels j and k):

$$U_{i\ldots} = k_0 + k_1 u_i + k_2 u_. + k_3 u_. + \ldots, \qquad (2)$$

where all terms are as previously defined, except for u.., which denote averages. Equation (2) tells us that the conditional mean utility estimates for each attribute level measure the utility associated with the each level of attribute 1 up to a positive linear transformation. That is, the conditional (marginal) means provide interval scale estimates of the utilities associated with the attribute levels. Similar expressions hold for each attribute, but the intercepts and slopes of the expressions for each attribute will differ. For example, if we collect terms and rearrange equation (2), we would have

$$U_{i\ldots} = \left(k_0 + k_2 u + k_3 u\right) + k_1 u_i + \ldots, \quad (3)$$

Equation 3 makes it easy to see that the intercept (the origin of the scale) and the slope (the unit of measurement of the scale) will differ for each attribute, and hence cannot be directly compared.

Now consider the fact that the term $k_1 u_i$ is a composite term that captures both the underlying scale position of the attribute and the importance of the attribute in the decision. That is, an attribute can impact the decision in the same way across all its levels (i.e., constant weight), or its weight can differ systematically across the levels. Alternatively, weights can be unitised or the same for all attributes, with utility values differing across levels. Hence, the “effect” (or lack thereof) of an attribute across its level can be due to a large (small) weight relative to other attributes, or can be due to large (small) differences in underlying utility values associated with the levels, or some combination of the two. It is not possible to determine which of these cases applies in a choice experiment without additional information that is not available to the researcher. Thus, we say that weight and scale value are confounded, implying that interdimensional utility comparisons reflect some composite of weight and scale value (position of attribute levels on the latent utility scale). In turn, this poses issues for generalising choice models because the results may be “level dependent”. That is, the effects one obtains in a choice model depend on the levels one chose to vary in the experiment, and it may well be that a different set of levels, or a different range of levels, or even a different (sub) set of attributes will yield different results (see for example Ohler, at al 2000 [1]; Louviere and Islam 2004 [2]). Thus, strictly speaking, conclusions about attribute effects should be qualified to be “relative” and not absolute, with stronger conclusions reserved for results that can be shown to generalise across different levels and subsets of attributes.

The preceding discussion reinforces the earlier point that to identify relative importance of each attribute one needs to measure them on a common, and therefore comparable, scale. One purpose of this paper is to propose three methods that allow one to make such comparisons: 1) partial log likelihood analysis; 2) calculation of marginal rates of substitution; and 3) calculation of Hicksian welfare measures. We demonstrate each method in an empirical application and compare them to the relative importance of attributes elicited from direct ranking and rating questions.

A second purpose of the paper is to propose a way to enhance the statistical efficiency of the parameter estimates in choice experiments. We propose a method that we term "Best Worst Choices" (BWC), to increase the statistical efficiency of random utility choice models by obtaining extra preference information from respondents. This involves asking respondents to choose their most and least preferred options in each choice set to produce a semi-order. A complete order can be obtained by successively repeating BWC questions for smaller subsets of the choice set until all options are ordered. The ordered information can be used to expand the amount of usable choice data, thereby increasing the statistical efficiency of the estimates. We demonstrate the use of BWC, and compare it to the results of a standard, single-choice per choice set approach in an empirical application.

The rest of the paper is organised as follows. In sections 2 and 3 we outline the methods used to investigate the relative importance of attributes and present the BWC method respectively. In section 4 we illustrate these methods in an empirical application. We discuss the results of the empirical illustrations and the methods in general in section 5 and section 6 concludes.

2. Methods to investigate relative importance

Here we present three methods that place the attributes on a common and therefore commensurable scale.

Partial LL

One way to compare the relative 'weight' or 'importance' of the attributes of the product or program to the choice at hand is to investigate the explanatory power of each attribute, or indeed each attribute level, by calculating the amount that each attribute contributes to the overall log likelihood of the choice model [3]. This involves systematically re-estimating the choice model omitting each attribute (or attribute level) one at a time and recording the associated log likelihood. The difference between the log likelihood of the full model and the log likelihood of omitting a particular attribute (level) can be interpreted as the partial effect of that attribute (level) on the choice responses if the design on which the analysis is based is orthogonal. Thus, attributes that are more 'important' to the decision at hand will contribute a larger amount to the total log likelihood, as indicated by their partial log likelihood.

This is analogous to calculating the partial r-square associated with each attribute in traditional rating and ranking tasks used in conjoint analysis (Louviere 1988). It also is related to statistical tests for the additional explanatory power of variables that are included/excluded from choice models. That is, the comparisons of interest are nested because the attributes can be included or excluded (a model selection problem), and the contribution of each attribute is simply (two times) the difference in the model log-likelihood for the attribute (with all its levels) in the model and out of the model. As the "two times" operator is a constant, it drops out, and one can simply compare the total log-likelihood difference associated with each attribute across all its levels.

Marginal rate of substitution

The marginal rate of substitution (MRS) can be used to investigate the rate at which individuals trade off one attribute for another. As such, this approach can be used to compare the relative importance of each attribute where each attribute is measured with respect to a common numeraire attribute; for example time or price. This places the attribute utility differences or particular attribute effects (e.g., a linear specification for waiting time in a clinic) on a common scale like minutes or dollars, thereby facilitating a comparison of the relative importance of each attribute. If the numeraire is price, we refer to this as calculating the 'implicit price' of each attribute.

Using the indirect utility function (IUF) estimated from a choice model and holding utility constant, following standard consumer theory the MRS is calculated by partially differentiating the IUF with respect to the first attribute of interest and then with respect to the second attribute of interest and taking the ratio

$$MRS_{X_1,X_2} = \frac{\partial V / \partial X_1}{\partial V / \partial X_2} \tag{4}$$

where V is the IUF estimated from the choice experiment and X_1 and X_2 are attributes of the good or service. The numerator is interpreted as the marginal utility of attribute 1 and the denominator is the marginal utility of attribute 2. Where time or price is used, the denominator would denote the marginal disutility of time or price.

Until now DCEs reported in the health economics literature have estimated linearly additive IUF of the form

$$V_j = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n \tag{5}$$

In this case the MRS between two attributes is simply the ratio of the coefficients on those attributes

$$MRS_{X1,X2} = \frac{\partial V / \partial X_1}{\partial V / \partial X_2} = \frac{\beta_1}{\beta_2} \tag{6}$$

where β_1 and β_2 are the estimated coefficients on X_1 and X_2 respectively; again where X_2 could be time or price.

However, utility need not be linearly additive [4]. In the empirical illustration included in this paper we estimate a non linear IUF of the form

$$V_j = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \beta_{12} X_1 X_2 + ... + \beta_{1n} X_1 X_n$$
$$+ .... + \beta_{n1} X_n X_1 + ... + \beta_{nn-1} X_n X_{n-1} \quad (7)$$

Equation (7) includes all main effects and all two way interactions. In such cases the MRS is no longer simply the ratio of the estimated coefficients on two attributes because attributes now enter utility both linearly and multiplicatively.

For example, when calculating the marginal utility of X_1 we must partially differentiate equation (7) with respect to X_1 including both the main effect of X_1 and where X_1 is interacted with other attributes.

In fact, as will be discussed below, in addition to including all two way interactions as described in equation (7) in our empirical illustration, we also include a quadratic term on one of the main effects, which again must be taken into account when calculating the MRS.

For completeness, we also note that in the case of qualitative attributes, we are interested in difference in the utility of two attribute levels. That is, if the provider of care can be doctor, nurse, paramedic, we are interested in the value of the difference – say, between doctor and nurse – but this difference is a single value, not a rate. To compare these differences we need a different concept, the Hicksian compensating variation that we now discuss.

Hicksian CV

The Hicksian compensating variation also can be used to measure the relative importance of each attribute on a common monetary metric by calculating the willingness to pay or to accept compensation for a change in a given attribute.

The importance of calculating theoretically consistent welfare measures from DCEs, and a method of calculating the Hicksian compensating variation in the context of a discrete choice random utility model has been discussed by Lancsar (2002) [5] and

Lancsar and Savage (2004a, 2004b) [6, 7]. Such welfare measures are calculated by using the estimated utility parameters in the following expression

$$CV = -\frac{1}{\lambda}\left[\ln\sum_{j=1}^{J} e^{V_j^0} - \ln\sum_{j=1}^{J} e^{V_j^1}\right] \quad (8)$$

where λ is the marginal utility of income, V_j^0 and V_j^1 are the value of the IUF for each choice option j before and after the quality change respectively, and J is the number of options in the choice set. Thus the CV essentially provides the change in the expected utility due to a change in the attribute(s) weighted by the marginal utility of income. It also takes account of the probability of choosing each alternative in the choice set. In general, it also has the advantage that one can now define the monetary equivalent associated with differences in two qualitative attribute levels. By extension, therefore, one can calculate the "total CV" associated with any attribute as the sum of the successive CV quantities associated with differences in the L and L-1 levels + the CV for the difference in the L-1 and L-2 levels + … + the difference in the L-(L-1) and L-(L-2) levels. This sum provides a metric for the total CV associated with all successive comparisons across levels.

In the empirical application reported below, we were interested in the CV associated with each level, rather than a change in levels, so we calculated the CV for a move from just including ASCs in the IUF to including ASCs plus each attribute level one at a time. Within the context of effects codes, the ASC is the expected utility when all attributes are equal to zero [8].

3. BWC

BWC is a way to obtain extra preference information by asking respondents to choose their most and least preferred options in each choice set, and from successively smaller subsets of each choice set. Specifically, for a choice set of size J, individuals can be asked to make up to J-1 choices. So in addition to the traditional approach of asking respondents to choose their most preferred alternative from the choice set, individuals also can be asked to choose their least preferred alternative. These two choices provide J-1 + J-2 choices. For example, if J=4, and best=a and worst=d, one knows a>b, a>c, a>d and b>d, c>d. Thus, one can expand the data to include these additional pairs. One also can ask individuals to choose their most and least preferred options from the remaining subsets of options to obtain a complete ranking of the choice options. The complete ranking then can be expanded into additional choice sets, although all the statistical information is contained in the full set of paired choices. Expanding the available data in this way increases the total number of observations several-fold, which in turn increases the statistical efficiency of the estimates. The theoretical basis for these expansions is the Luce and Suppes (1965) [9] Ranking Theorem, a well-known way of expanding a complete ranking. However, even semi-orders contain additional choice information; hence, merely asking for the most and least preferred options in each set dramatically increases the amount of available choice information. However, using only best and worst choices results in biased estimates of model intercepts (or alternative-specific constants) as noted by Ben-Akiva and Lerman (1985) [10]. That is, with only best and worst choices, the probability of choosing an option is not independent of the probability that it is available to be chosen (e.g., in the example where J=4, if a is best, it will “appear” more often in the expanded sets than the other options). The model parameters are not affected by this; hence, if one’s objective is policy analysis and/or drawing inferences from the model estimates about impacts of attributes/levels, the bias in the intercepts does not matter. If one wants to use the models to predict choice probabilities, however, one needs to adjust the model intercepts as noted by Ben-Akiva and Lerman [10].

4. Empirical application

4.1 Data

We now demonstrate the proposed methods using data from a DCE designed to investigate preferences for treatment of cardiac arrest occurring in a public place. A sample of 64 people drawn from the general public in Calgary, Alberta, Canada, were presented with 16 choice sets. Each choice set offered 4 treatment options described by 5 attributes (2 with 2 levels, 2 with 4 levels and 1 with 8 levels as described in Table 1). An optimally efficient design was used to construct 512 scenarios that were split into 32 versions of 16 choice sets by randomly assigning scenarios to versions without replacement [11]. The design allowed independent estimation of all main effects and two way interactions, thereby allowing us to estimate a non-linear indirect utility function, which to our knowledge is novel in the health economics literature.

Insert Table 1: Attributes and levels

The first option in every choice set was to stay with the status quo treatment for cardiac arrest occurring in a public place; namely to wait for the ambulance to arrive. The remaining 3 treatment options per choice set described different versions of what is termed 'public access defibrillation'; that is, having an automated external defibrillator available in a number of public places to be used while waiting for an ambulance to arrive.

In each choice set respondents were asked three questions: 1) to choose the best treatment option; 2) to choose the worst treatment option; and 3) to choose the best of the remaining 2 options. Thus, in addition to the standard discrete choice regarding the most preferred alternative, a complete ranking of the 4 alternatives were obtained in each of the 16 choice sets. To estimate the BWC model, we used the implied ranking to calculate weights associated with each rank position (i.e., a rank of 1 = 8, 2 = 4, 3 = 2 and 4 = 1). These weights reflect the fact that there are 2^4 possible choice sets of 4 choice options, and the individual has to choose one of them, even if that is to stay with the status quo. Thus, if the individual chooses consistently with the implied rank order of the options, the first ranked would be chosen 8 times in the 16 sets, the second ranked 4 times, etc. Thus, the weights simply represent a convenient way of organising the data. Naturally, one can expand the ordering to all of the implied pairs or all of the implied choice sets instead of using weights. It also is worth noting that in cases where individuals do not have to make a choice, one should choose a different set of weights to reflect the fact that the single element choice sets (one option) provide no choice information. In the case of 4 options, these revised weights would be 7, 3, 1, 0.

4.2 Results

BWC and DCM

We initially estimated a model based on the BWC weights and a traditional first-choice discrete choice model (DCM) with main effects only. All attributes were effects coded in order to visualise the results and infer a more parsimonious reduced form utility expression if possible [8]. These results suggested that the effects of the survival levels were non-linear; hence, we estimated models that specified survival with linear and quadratic effects, as discussed below. We mean-centered the eight level price attribute, and coded the remaining three attributes with effects codes.

The estimation results for the BWC model including main effects only (Model 1) and main effects + all two-way interactions (Model 2) are in Table 2.

Insert Table 2: BWC model

All main effects were statistically significant at the 1 percent level except for location of care in the main effects only model (Model 1). The Model 2 results suggest that few interactions are significant. Although location is not significant individually it is significant when interacted with provider and form of payment.

The results of the DCM, using only the first choice information parallel that for the BWC analysis with a main effects only model (Model 3) and a model for main effects + all two-way interactions (Model 4); these results are in Table 3.

Insert Table 3: DCM

Again, all main effects are statistically significant at the 1 percent level except for location of care in the linearly additive model (Model 3), with few interactions significant in the non-linear model (Model 4). In Model 4 the interaction of survival and provider and survival and from of payment are statistically significant.

Comparing models, the DCM had larger coefficients in absolute terms, although they were proportional as shown in the scatter plots of the two sets of estimates for Models 1 and 3 and Models 2 and 4 in Figures 1 and 2 respectively. The BWC models exhibited smaller standard errors on every effect, reflecting the gains in statistical efficiency. The DCM exhibit slightly larger McFadden Rho-squares and smaller log likelihoods, which reflects the fact that expanding the choices leads to more variability to be explained by the BWC models, in turn resulting in lower fits.

Insert Figure 1: BWC v DCM Coefficients: Model 1 v Model 3

Insert Figure 2: BWC v DCM Coefficients: Model 2 v Model 4

Partial LL

The results of the partial log likelihood analysis undertaken for the BWC model including the main effects + all two-way interactions are in Table 4. These results are based on 25 model estimations in which we systematically included/removed a particular attribute level from the model. The log likelihood values associated with each model are in the second column of Table 4. For each attribute level, the change in log likelihood (column 3) is calculated by subtracting the reduced model log likelihood from the full model log likelihood. The relative effect is calculated as the percentage change in log likelihood in column 4, the cumulative effect is in column 5 and the implied ranking of attribute level effects is in column 6.

Insert Table 4: Partial LL Analysis

Not surprisingly, high ranking attributes also were significantly different from zero in Model 4. Interestingly, location is not 'important' individually, in the sense that it has a negligible impact on the log likelihood, but is relatively 'important' when interacted with provider and form of payment, which have the sixth and seventh largest impact on the log likelihood.

Survival accounted for 68 percent of the log likelihood, including price, provider and form of payment collectively accounts for 96 percent of the log likelihood. Adding the interactions increased the log likelihood marginally.

As can be seen from table 4, the variables that were not significant in the BWC model (the location levels and most interactions) are included in the partial log likelihood analysis because this lack of significance is taken into account in the estimation of the partial log likelihoods. However, these (non-significant) attributes are excluded from the remaining examination of relative importance because the BWC model suggests that the location levels and most interactions are not significantly different to zero. Thus, we eliminate these from further consideration.

Marginal rate of substitution

The marginal rate of substitution between price and all other attributes that were statistically significant in Model 3 are presented in Table 5.

Insert Table 5: MRS

As discussed in Section 2, when calculating these MRS we took account of the non-linear IUF for both non linear main effects and also for significant interaction terms. By way of example, the attribute 'chance of survival' was decomposed into a linear and a quadratic term, and the interaction of these terms with the 'provider' attribute also were significant. Thus, the MRS between price and survival is no longer a ratio of the coefficients on the main effects. Instead, the MRS between survival and price is obtained by partially differentiating the IUF first with respect to survival and then with respect to price to give

$$MRS_{S,P} = \frac{\beta_{S\_lin} + (2 * \beta_{S\_quad}) + (2 * \beta_{Sq\_pr})}{\beta_P}$$

where β_{S_lin} is the coefficient on the linear survival term, β_{S_quad} is the coefficient on the quadratic survival term and β_{Sq_pr} is the coefficient on the interaction between quadratic survival and provider.

The location level, library, was only significant via its interaction with attributes form of payment and provider; hence, these terms were included in the calculation of the MRS.

Hicksian CV

The results of the welfare analysis using equation (5) are in Table 6. These welfare measures were calculated by taking account of the non-linear nature of the estimated IUF. That is the IUFs, the Vs, in equation (5) included significant interaction and non-linear effects.

Insert Table 6: Welfare measures

Comparison of the relative importance of attributes

The results of the various methods outlined above to investigate the relative importance of the attributes of the BWC model are compared to each other and to the results of a direct ranking and rating question in Table 7.

Table 7: Ranking of relative importance of attributes across methods

The ranking and rating results presented in the last two columns of Table 7 were calculated by summing and taking the average of responses to direct ranking and rating questions in the survey and ranked in order of magnitude.

The comparison in Table 7 indicates that the chance of survival attribute was consistently ranked the most important attribute across all methods and location was consistently the least important. This may reflect the view that individuals have fully formed preferences about the attributes they do and do not like but there is less certainty around preferences for attributes that fall in between.

5. Discussion

BWC model

The BWC method is relatively easy for respondents as it involves making simple discrete choices. It also provides an implied ranking without asking respondents to directly rank options. The additional implied choice set responses increases statistical efficiency as shown by the smaller standard errors relative to those estimated in the standard DCM.

Previous work suggests that respondents may prefer what they know best in terms of the status quo (Salkeld et al, 2000). Thus, an advantage of BWC may be that by having respondents make more than one choice per choice set, BWC encourages respondents to consider all options and trade off levels of attributes across alternatives, potentially reducing status quo bias or indeed the choice of the neither option if one is included. This warrants further research.

The utility functions for both BWC and DCM models were specified as non-linear expressions (they include interactions and non-linear attribute effects). Although few interactions were significant, this may be due to the small initial sample size in the study that we used to illustrate the methods.

Relative importance

We proposed, discussed and illustrated three ways to measure the relative importance of attributes and/or levels. In the case of the MRS and Hicksian CV, as previously noted, it is the absolute value of the results that matters because attributes can have a positive *or negative* impact on utility or WTP and it is the size of this impact rather than the direction that is of interest when looking at the ‘weight’ or importance of each attribute.

Table 7 highlights that while there are similarities in the rankings across the methods, there also are differences. Indeed reviews of the concept of attribute ‘importance’ or ‘weight’ undertaken by Shanteau (1980) [12] and Louviere and Islam (2004) [2] reveal little agreement either in the definition or the approach to measuring ‘importance’, with the result that different ways of measuring attribute importance can lead to different conclusions.

While, which method is most appropriate to investigate the issue of relative ‘weight’ will in part depend on the purpose of the study or the research questions to be addressed, there are advantages and disadvantages of each.

If a continuous attribute is included in a DCE, the MRS between this attribute as numeraire and all other attributes provides a way to measure the relative importance of attributes. However, calculations become more involved if the utility specification is non-linear and/or non-additive, as we demonstrated above. Of course, one also may want to measure the relative impact of the attribute used as the common base (such as price or time), which cannot be accomplished using the MRS.

The Hicksian compensating variation approach provides a viable alternative to also measure the relative importance of numeraire attributes like price or time as the marginal utility of income can be used to convert the impact of other attributes into monetary terms, rather than using one of the attributes of the choice model (such as price or time). It would make sense to use this method if calculating welfare measures were an objective of the study independent of investigating relative importance.

If one only wishes to measure the overall effects of the attributes relative to one another, and one does not want to derive policy measures like the compensating variation, the partial log likelihood approach provides a way to do this. The appeal of this approach lies in the fact that it does not require one of the attributes to be used as a common base. It also measures the impact of each attribute across its levels, in a simple and intuitive way, by estimating the relative contribution of each level to the explanatory power of the model.

Current work is focused on investigating the relative importance of attributes by estimating their effect on the probability of choosing a particular alternative.

6. Conclusion

This paper discussed the fact that despite common practice, the relative importance of attributes in a choice experiment cannot be inferred directly from the estimated coefficients due to confounds between the 'weight' or importance of the attribute in the decision and the utility scale on which the attribute levels lie. We presented three methods that can be used to measure attributes on common and comparable scales: partial log likelihood analysis; MRS, and welfare measures. The use of partial log likelihood analysis in particular is novel in the health economics literature.

We also presented a new method that we termed "Best Worst Choices" that involves asking additional choice questions about options in each choice set. The resulting answers allow one to expand the available choice data, which increases the statistical efficiency of choice models estimated from the resulting data.

In addition, we discussed and illustrated the estimation of a non linear IUF that included all of the two-way attribute interactions. To our knowledge, such a utility specification has not previously been considered in the health economics literature. We also demonstrate how to derive the MRS from such an IUF.

Tables

Table 1: Attributes and levels

Attributes	Levels
Chance of survival with treatment	• 6 out of 100 • 9 out of 100 • 12 out of 100 • 15 out of 100
Provider of care	• Trained responder • Non-trained responder
Location	• Shopping mall • Gym or other sports centre • Senior centre • Public library
Price	• $170 • $200 • $230 • $260 • $290 • $320 • $350 • $380
Method of payment	• Direct out of pocket payment • A one off increase in taxation

Table 2: BWC Analysis

	Model 1: BWC model, main effects only (MNL)		Model 2: BWC model, main effects + all 2-way interactions (MNL)	
Attribute	Coefficient	Standard Error	Coefficient	Standard Error
payform_OPP	0.08783***	0.01043	0.08966***	0.01073
payform_tax	-0.0878		-0.08966	
prov_nontrained	-0.09412***	0.01048	-0.09475***	0.01075
prov_trained	0.09412		0.09475	
loc_mall	0.03086	0.02429	0.02741	0.02471
loc_library	-0.01373	0.02427	-0.01455	0.02468
loc_gym	0.01589	0.02435	0.01721	0.02471
loc_seniorcentre	-0.03302		-0.03007	
s_lin	0.15421***	0.00566	0.15472***	0.00570
s_quad	-0.01775***	0.00321	-0.01696***	0.00323
pricemc	-0.00241***	0.0018	-0.00239***	0.00018
slin_p			-0.00005	0.00008
sq_p			0.00004	0.00004
p-f			-0.00013	0.00017
p-pr			-0.00025	0.00016
slin_f			-0.00737	0.00495
sq_f			0.00131	0.00267
slin_pr			0.00775	0.00538
sq_pr			0.0101***	0.00290
f_pr			-0.00692	0.01064
l1_p			0.00002	0.00029
l2_p			0.00001	0.00029
l3_p			0.00004	0.00028
slin-l1			0.00877	0.00928
sq_l1			-0.00717	0.00524
slin_l2			0.01457	0.00934
sq_l2			-0.002	0.00526
slin_l3			-0.00861	0.00922
sq_l3			0.001	0.00524
l1_f			0.0017	0.01787
l2_f			-0.03187*	0.01794
l3_f			0.01697	0.01789
l1_pr			-0.00661	0.01823
l2_pr			0.03973**	0.01830
l3_pr			-0.01976	0.01826
constant_PADA	-0.41030***	0.02310	-0.41158***	0.02312
constant_PADB	-0.25126***	0.02206	-0.25285***	0.02208
constant_PADC	-0.45599***	0.02342	-0.45804***	0.02345
Log likelihood	-19928.681		-19909.337	
McFadden R^2	0.02738		0.02833	

*** significant at 1%, ** significant at 5%, * significant at 10%

McFadden's R^2 is defined as 1 – (LL/LL_0), where LL is the value of the (simulated) log-likelihood function evaluated at the estimated parameters while LL_0 is the value of the log-likelihood function for a base model that only contains a non-random alternative-specific intercepts.

Table 3: DCM Analysis

	Model 3: DCM, main effects only (MNL)		Model 4: DCM, main effects + all 2-way interactions (MNL)	
Attribute	Coefficient	Standard Error	Coefficient	Standard Error
payform_OPP	0.15046***	0.04373	0.20887***	0.05581
payform_tax	-0.15046		-0.20887	
prov_nontrained	-0.16994***	0.04478	-0.21326***	0.05565
prov_trained	0.16994		0.21326	
loc_mall	0.0231	0.09885	0.02926	0.11332
loc_library	-0.00265	0.09879	-0.10667	0.12101
loc_gym	0.00532	0.09924	0.05352	0.11136
loc_seniorcentre	-0.02577		0.02389	
s_lin	0.39356***	0.02891	0.42246***	0.03255
s_quad	-0.05243***	0.01454	-0.05877***	0.01557
pricemc	-0.00377***	0.00076	-0.00358***	0.00087
slin_p			-0.00025	0.00042
sq_p			0	0.00020
p-f			-0.00036	0.00071
p-pr			-0.00027	0.00069
slin_f			-0.06022**	0.02822
sq_f			0.00876	0.01318
slin_pr			0.05828**	0.02865
sq_pr			0.01132	0.01387
f_pr			-0.04574	0.04842
l1_p			0.00013	0.00123
l2_p			-0.00033	0.00126
l3_p			0.00112	0.00122
slin-l1			0.00603	0.04975
sq_l1			-0.0061	0.02462
slin_l2			0.10571	0.05646
sq_l2			-0.03708	0.02647
slin_l3			-0.04082	0.04790
sq_l3			0.01336	0.02432
l1_f			-0.00572	0.07652
l2_f			-0.0035	0.07731
l3_f			-0.05249	0.07619
l1_pr			0.01261	0.07792
l2_pr			0.08449	0.07843
l3_pr			-0.0677	0.07838
constant_PADA	-1.11445***	0.10279	-1.16882***	0.10602
constant_PADB	-0.55068***	0.08688	-0.59842***	0.09037
constant_PADC	-0.97912***	0.09865	-1.03184***	0.10199
Log likelihood	-1168.2208		-1155.33009	
McFadden R^2	0.11269		0.12248	

*** significant at 1%, ** significant at 5%, * significant at 10%

McFadden's R^2 is defined as $1 - (LL/LL_0)$, where LL is the value of the (simulated) log-likelihood function evaluated at the estimated parameters while LL_0 is the value of the log-likelihood function for a base model that only contains a non-random alternative-specific intercepts.

Figure 1: BWC v DCM coefficients: Model 1 v Model 3

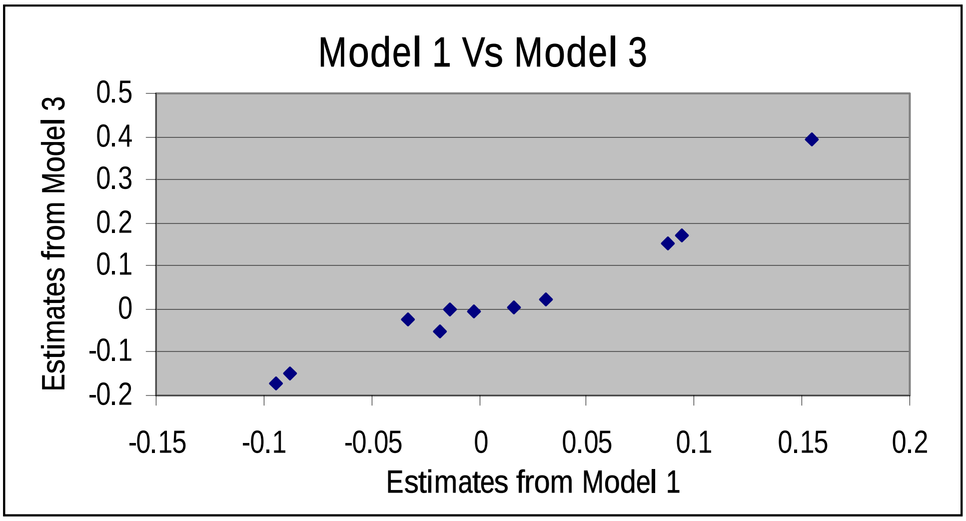

Figure 2: BWC v DCM coefficients: Model 2 v Model 4

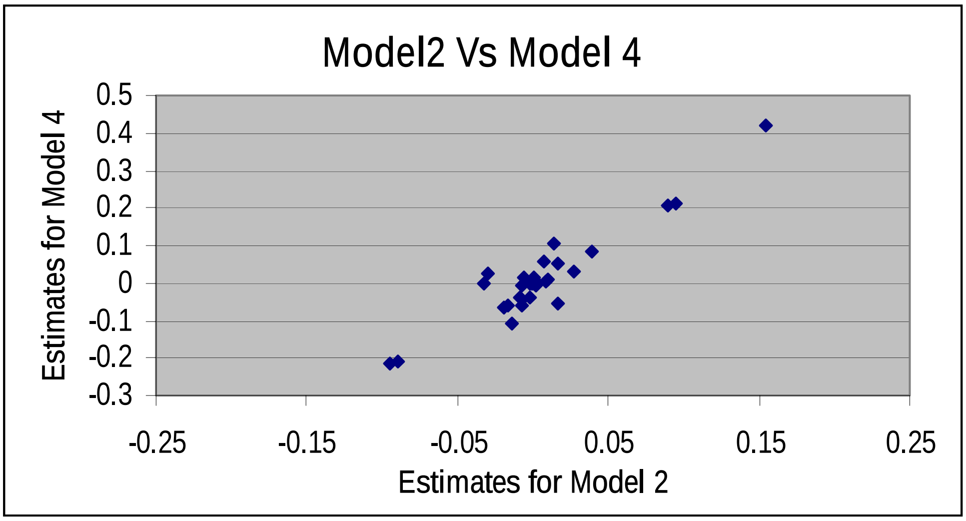

Table 4: Partial LL Analysis

Attribute level excluded from the analysis	Log likelihood	Partial effect - change in log likelihood	Relative Effect - % sum of change in log likelihood	Cumulative %	Attribute importance rank
None (full model)	-19909.3367				
no surv (lin+quad)*	-20303.22824	-393.89154	0.686	0.6855715	1
no price*	-19995.59421	-86.25751	0.150	0.8357034	2
no provider*	-19948.36856	-39.03186	0.067935	0.9036387	3
no form*	-19944.37765	-35.04095	0.060989	0.9646277	4
no slin_prov or sq_prov*	-19917.36578	-8.02908	0.013975	0.9786024	5
no loc2_prov*	-19911.69248	-2.35578	0.004100	0.9827027	6
no loc2_form*	-19910.91328	-1.57658	0.003	0.9854467	7
no slin_loc2 or sq_loc2	-19910.56873	-1.23203	0.002	0.9875911	8
no slin_loc1 or sq_loc1	-19910.55539	-1.21869	0.002121	0.9897122	9
no price_provider	-19910.53469	-1.19799	0.002085	0.9917973	10
no slin_form or sq_form	-19910.46734	-1.13064	0.001968	0.9937652	11
no location1	-19909.95108	-0.61438	0.001069	0.9948345	12
no loc3_prov	-19909.92285	-0.58615	0.001020	0.9958547	13
no slin_price or sq_price	-19909.83505	-0.49835	0.000867	0.9967221	14
no loc3_form	-19909.78709	-0.45039	0.001	0.9975060	15
no slin_loc3 or sq_loc3	-19909.773	-0.4363	0.001	0.9982654	16
no price_form	-19909.62219	-0.28549	0.000497	0.9987623	17
no location3	-19909.57893	-0.24223	0.000422	0.9991839	18
no form_provider	-19909.54825	-0.21155	0.000368	0.9995521	19
no location2	-19909.51053	-0.17383	0.000303	0.9998547	20
no loc1_prov	-19909.40242	-0.06572	0.000	0.9999691	21
no loc3_price	-19909.34773	-0.01103	0.000019	0.9999883	22
no loc1_form	-19909.34123	-0.00453	0.000	0.9999961	23
no loc1_price	-19909.33873	-0.00203	0.000004	0.9999997	24
no loc2_price	-19909.33689	-0.00019	0.000000	1.0000000	25

* Significant in BWC model

Table 5: MRS

Attribute	MRS with P	Absolute value of MRS with P	Attribute importance rank
Chance of survival	-59	59	1
Out of pocket payment	-24	24	2
Non trained provider	23	23	3
Library	-3	3	4

Table 6: Welfare measures

Attribute	CV *	Absolute value	Attribute importance rank
Chance of survival	-43	43	1
Non trained provider	-27	27	2
Tax	25	25	3
OPP	-16	16	4
Trained provider	8	8	5
Library	4	4	6

* For a move from the grand mean to the grand mean plus the attribute of interest

Table 7: Ranking of relative importance of attributes across methods

Attribute	Partial LL	MRS	Welfare measure	Ranking	Rating
Main effects					
Survival	1	1	1	1	1
Price	2			4	3
Form of payment	4	2		2	2
• payform_OPP			4		
• payform_tax			3		
Provider	3	3		3	4
• Non trained			2		
• Trained			5		
Location				5	5
• Mall					
• Library		4	6		
• Gym					
• Senior centre					
Interactions*					
survival x provider	5				
location 2 x form	7				
location 2 x provider	6				

*The effect of significant interactions are included in the effect of the main effect in the calculation of the MRS and welfare measures

References

1. Ohler T, Le A, Louviere JJ, Swait J. Attribute Range Effects in Binary Response Tasks. *Marketing Letters* 2000; **11**: 249-260.
2. Louviere JJ, Islam T. *A Comparison of Importance Weights/Measures Derived from Choice-Based Conjoint, Constant Sum Scales and Best-Worst Scaling*. University of Technology, Sydney 2004.
3. Crouch GI, Louviere JJ. The Determinants of Conventions Site Selection: A logistic Choice Model from Experimental Data. *Journal of Travel Research* 2004; **43**: 118-130.
4. Viney R, Lancsar E, Louviere J. Discrete Choice Experiments to Measure Consumer Preferences for Health and Healthcare. *Expert Review of Pharmacoeconomics and Outcomes Research* 2002; **2**: 319-26.
5. Lancsar E. *Deriving welfare measures from stated preference discrete choice modelling experiments. CHERE Discussion Paper No 48*. Centre for Health Economics Research and Evaluation, University of Technology, Sydney 48. 2002.
6. Lancsar E, Savage E. Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory. *Health Economics Letters* 2004; **13**: 901-907.
7. Lancsar E, Savage E. Deriving welfare measures from discrete choice experiments: a response to Ryan and Santos Silva. *Health Economics Letters* 2004; **13**: 919-924.
8. Louviere JJ, Hensher, D.A, Swait, J.D. *Stated Choice Methods Analysis and Application*. Cambridge University Press: Cambridge, 2000.
9. Luce RD, Suppes P, editors. *Preference, Utility and Subjective Probability*, 1965.
10. Ben - Akiva ME, Lerman S. *Discrete Choice Analysis: Theory and Applications to Travel Demand*. The MIT Press: Cambridge, 1985.
11. Burgess L, Street DJ. Optimal designs for choice experiments with asymmetric attributes. *Journal of Statistical Planning and Inference* 2004; **In press**.
12. Shanteau J. *The Concept of Weight in Judgement and Decision Making - A Review and Some Unifying Proposals*. Centre for Research and Judgement and Policy, Institute of Behavioral Research, University of Colorado, Boulder, Colorado report no 228. 1980.