Journal of Choice Modelling

Journal of Choice Modelling, 1(1), 2008, 128-164
www.jocm.org.uk

Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information

Jordan J. Louviere[1,*] Deborah Street[2,†] Leonie Burgess[2,‡]
Nada Wasi[1,§] Towhidul Islam[1,3,¶] Anthony A. J. Marley[4,‖]

[1] School of Marketing & Centre for the Study of Choice (CenSoC), University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia

[2] Department of Mathematical Sciences & Centre for the Study of Choice (CenSoC), University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia

[3] Department of Marketing and Consumer Studies, University of Guelph, Ontario, Canada

[4] Department of Psychology, University of Victoria, PO Box 3050 STN CSC, Victoria, British Columbia, Canada

Received 29 September 2007, revised version received 6 April 2008, accepted 28 July 2008

Abstract

We show how to combine statistically efficient ways to design discrete choice experiments based on random utility theory with new ways of collecting additional information that can be used to expand the amount of available choice information for modeling the choices of individual decision makers. Here we limit ourselves to problems involving generic choice options and linear and additive indirect utility functions, but the approach potentially can be extended to include choice problems with non-additive utility functions and non-generic/labeled options/attributes. The paper provides several simulated examples, a small empirical example to demonstrate proof of concept, and a larger empirical example based on many experimental conditions and large samples that demonstrates that the individual models capture virtually all the variance in aggregate first choices traditionally modeled in discrete choice experiments.

Keywords: Choice Models, Econometric Models, Discrete Choice Experiments

*Corresponding author, T: +61 2 9514 3522,F: +61 2 9514 3535, jordan.louviere@uts.edu.au
†T: +61 2 9514 2251,F: +61 2 9514 2260, deborah.street@uts.edu.au
‡T: +61 2 9514 1965,F: +61 2 9514 2260, leonie.burgess@uts.edu.au
§T: +61 2 9514 3203,F: +61 2 9514 3535, nada.wasi@uts.edu.au
¶T: +1 519 824 4120 ext. 53835 ,F: +1 519 823 1964, islam@uoguelph.ca
‖T: +1 250 472 2067,F: +1 250 721 8929, ajmarley@uvic.ca

ISSN 1755-5345

1 Introduction

From time-to-time since Thurstone (1927) developed random utility theory (RUT), attempts have been made to develop ways to estimate models for individual decision makers for problems of typical size seen in survey applications. For example, Carson et al. (1993) noted that four choice options and eight attributes were typical of surveys that implement the types of discrete choice experiments (DCEs) proposed by Louviere and Woodworth (1983). It is fair to say that most researchers believe that a) one would need to observe many choice sets for each person in a survey to insure that each *individual-level discrete choice model* would successfully converge and yield reliable estimates, and b) the number of choice sets required would be too large for typical survey applications (see, e.g., Chapman, 1984). Thus, at a minimum, most researchers probably would want to have a complete or at least partial ranking across the alternatives in each choice set in a DCE, and sufficient numbers of choice sets to insure identification of effects for a person.

An extreme example of using "many choice sets" to model one person was provided by Pihlens and Louviere (2004), who administered several thousand pairs of options to estimate a model of the color choices for one person in a purely academic study. As we discuss below, some researchers have developed or applied one or more so-called "adaptive" ways to estimate models for individuals, such as adaptive polyhedral methods that are used with paired comparisons (e.g., Toubia et al., 2003, 2004). Other approaches to estimating model parameters for single persons are based on continuous or finite distribution models that we discuss later. These approaches model individuals indirectly (e.g., Revelt and Train, 1998), whereas our focus in this paper is on modeling individuals directly, not modeling them indirectly based on assumptions about preference distributions across samples of people.

Approaches also have been proposed similar to those of Toubia et al. (2003, 2004) broadly termed "adaptive conjoint or choice methods" (e.g., Richardson, 2002; Shinghal and Fowkes, 2002). All adaptive methods of which we are aware select future choice sets/options to show respondents based on prior realizations of the dependent variable (choices, ratings, etc), which raises issues of selection bias (Heckman, 1979, 1990). Selection bias is defined as sampling based on the dependent variable, and adaptive methods currently lack formal proofs that they are exempt from selection bias. So, we do not discuss adaptive methods further.

There also are other potentially serious issues associated with individualization of choice experiments that is a shared feature of all adaptive methods. For example, many researchers seem unaware that there can be (and often are) version effects in choice experiments associated with particular sets of choice or comparison sets assigned to individuals. For example, some sets may have more high/low prices or tariffs that can lead to: a) lower/higher choices of some options, resulting in differences in alternative specific constants; b) lower/higher choices of options with particular attribute levels, resulting in "apparent" differences in sensitivity to attributes, such as higher/lower price sensitivity; and c) lower/higher choice variability within and between individuals, leading to parameter differences due to scale difference. Unfortunately, if one completely individualizes a stated preference experiment, one confounds such differences with differences in individuals, which can lead to biased estimates as well as incorrect attribution of effects to one source when in fact they are due to one or more other sources.

Some researchers obtained a full rank-ordering of the alternatives in particular choice sets, such as Ben-Akiva et al. (1991); Hausman and Rudd (1988) (see also Louviere et al., 2000, pp 25-33; Train, 2003, pp 160-168). Typically, this ranking was for one or a small number of sets, and we are unaware of applications where a full or partial ranking of the alternatives in each choice set was combined with a design for multiple choice sets to provide sufficient choice observations to estimate models for individuals, but naturally this can be done. Many researchers outside of psychology seem to think that asking an individual to rank a set of options is equivalent to asking her to make repeated preference choices from a set of options, such as most preferred, least preferred, next most preferred, etc. However, we note that this may/may not be true in any empirical application, and in any case, the process that any particular individual uses to order options may differ with two or more types of task formats. Which is "correct" in any particular case is an empirical question. We rely on repeated best and worst (most and least) choices because this seems to be easy for people, and as we later show, produces empirically reliable data.

Finally, we note that Horsky and Rao (1984) considered modeling individuals with full ranking responses based on all paired comparisons combined with attribute ratings, where the options ranked and rated were brands. Their approach is tangentially related to what we propose in this paper, corresponding to a case where one models only choice of objects (here, brands), using non-designed covariates to "explain" the choices (see also Severin et al., 2001). Our proposed approach does not use specific named choice options like brands, but instead generic options described by specifically designed attribute combinations. We are unaware of other theoretical or logical approaches to directly model an individual's choices that can be used in typical survey applications of discrete choice experiments.

To anticipate the approach we propose, we show how to model the choices of individual decision-makers by combining two recent developments into a single, integrated approach. These developments are 1) methods for constructing statistically efficient experiments to study and evaluate models of discrete choices (hereafter, Discrete Choice Experiments, or DCEs), such as those developed by Deborah Street and Leonie Burgess (see Street and Burgess, 2007, and references cited therein; see also Street et al., 2005, and Burgess et al., 2006); and 2) using most and least preferred/likely to choose/etc questions to easily obtain a full ranking of options, which in turn can be expanded into significant numbers of implied choices.

Our proposed approach produces response data that can be used with several logically and theoretically consistent ways to expand choice response data, such as rank-order explosion based on the Luce and Suppes (1965) Ranking Theorem, used by Chapman and Staelin (1982) to decompose preference rankings into a series of un-ranked and statistically independent choice observations. The rank-order explosion method allows one to more efficiently estimate a multinomial logit (vis, conditional logit model, CLM) model for a sample of individuals compared to using only first choice information. In contrast to prior applications of rank-order expansion, our proposed approach allows us to estimate a model for each person in a sample, not just a model for the sample as a whole. Later, we discuss several expansion methods and show empirically that they produce more or less the same model estimates "up to scale"; implying that our approach is compatible with several logical ways that researchers might choose to expand data and estimate models.

Any proposal of a way to model the choices of single individuals leads to two questions 1) why one would *want* to model individuals; and 2) what are the advantages of the approach proposed.

The answer to the first question is immediate. Researchers have long believed that individuals differ, i.e., there exists heterogeneity in tastes and error variances (or equivalently, scales) and have been interested in capturing individual effects. In theory, there exist statistical models that can incorporate individual's fixed effects into mean utility and/or variance *if* there is a sufficiently large number of observations per individual (see e.g., Greene, 2003, p.697).

That is a big *if* given that typically the number of observations per person is small for both revealed and stated preference data. Modern practice in discrete choice modeling relies on an analogue of random effects models. That is, it begins with an assumption about underlying unobserved taste heterogeneity in the population like "latent segments" (e.g., a latent class model) or a continuous distribution (e.g., a mixed logit model). Then, to find where in the distribution of tastes a particular individual resides, one estimates the mean of the distribution of the subpopulation, conditioning on a person's past choices. A similar approach is to use hierarchical Bayes (HB) estimation where information on the aggregate preference distribution is combined with individual's choices to calculate conditional estimates of individual consumers' preferences. Huber and Train (2001) compared individual level estimates from mixed logit and hierarchical Bayes, showing that they yield similar results. We call the foregoing approach "top-down" modeling.

The contributions of our paper are two fold: 1) we propose an efficient way to collect a large number of observations per person; and 2) we propose an alternative way to analyze the data that is simpler than current "top down" practices described above. The virtue of simplicity is that it allows less sophisticated applied researchers to learn and apply models that take individuals' heterogeneity into account.

First, we propose using optimally efficient DCE designs together with best-worst questions to obtain multiple observations per choice set per individual.

Compared with traditional "most-preferred" choice questions, best-worst provides extra information about individuals' preferences; and compared to traditional ranking questions (e.g., Ben-Akiva et al., 1994), places less cognitive burden on respondents (Hensher et al., 1999). Note that, even in a top-down approach, having a large number of observations per individual is essential. For instance, Train (2003, p.269) notes that the mean estimate of the conditional distribution can be considered to be the estimate of that person's parameter only as the number of choice situations that a person faces increases.

Second, because our approach provides many observations per individual, we propose that one can estimate individual parameters directly. This can be done in several ways using standard estimation software, which is simpler than estimating parameters of the conditional distribution. Our alternative approach obtains parameters for each person, which essentially gives the empirical distribution of sample preferences. If one wants to characterize these distributions, one can try to find appropriate distributions to fit them. Thus, we call this alternative process a "bottom-up" approach.

In theory, if one specifies correct distributions, and the number of observations per person is sufficiently large, "top-down" and "bottom-up" approaches should give the same results - for each individual's parameters and for parameters that characterize the sample distributions. If assumptions about preference distributions are incorrect, the inferences from top-down models will be biased and incorrect (e.g., Louviere et al., 2000, 2002; Louviere, 2004a,b; Louviere and Eagle, 2006).[1]

Thus, the purpose of this paper is to propose, describe, illustrate and discuss a relatively simple way to collect and model individual choices by combining statistically efficient designs for DCEs with repeated most and least preferred/likely choice questions about choice options in the designed choice sets. We consider *only* conditional logit models with strictly additive (i.e., main effects only) indirect utility expressions, but we note in the Discussion section that it is likely that this can be relaxed for some problems.

The rest of the paper is organized as follows: First we motivate several reasons why our approach "works". Then we briefly review efficient design of DCEs and discuss the role of optimally efficient experiments in producing efficient estimates of choice model parameters. Next we illustrate the proposed approach with two empirical examples to show proof of concept and demonstrate that the estimated models capture virtually all of the variability in traditional first choice data in traditional DCEs. Finally, we conclude by discussing unresolved issues and future research directions.

[1] Indeed, Magidson and Vermunt (2007, forthcoming) recently modified their commercial software to account for non-constant error variances in samples of individuals, demonstrating that one obtains biased estimates if these variances are not constant as commonly assumed in most "top-down" modeling approaches.

2 Efficient Designs for Discrete Choice Experiments

We focus on DCEs consistent with conditional logit models (CLMs) with indirect utility expressions restricted to "generic, main effects only", restrictions consistent with so-called "generic" choice experiments, such as those discussed in Louviere et al. (2000, pp. 119-120).

One typically compares the statistical efficiency of experimental designs by using some function of the variance-covariance matrix of the parameter estimates. Thus, one might choose to focus on minimizing some function of the variance of the parameter estimates, or one might focus on minimizing (a function of) the variance of the predictions. Whichever approach is adopted, the best designs are those that have the smallest possible value of the chosen function.

The most common optimality criterion for DCEs is D-optimality, which minimizes the generalized variance, and A-optimality, which minimizes the sum of the variances of the parameter estimates. If we denote the information matrix of a given design by C, then the variance-covariance matrix is given by C^{-1}. The D-optimal value of a design is the determinant of the variance-covariance matrix which is written $\det(C^{-1})$. The A-optimal value is $\mathrm{tr}(C^{-1})$, the trace of the variance-covariance matrix. Various authors discussed the construction of D-optimal designs including El-Helbawy and Bradley (1978), van Berkum (1987), Kuhfeld et al. (1994), Bunch et al. (1996), Sandor and Wedel (2001), Kanninen (2002) and Burgess and Street (2005), among others. More recently, Kessels et al. (2006) investigated the construction of designs that exhibit G-optimality, which minimizes the maximum prediction variance, and V-optimality, which minimizes the average prediction variance. These criteria are clearly closely related to D-optimality and A-optimality respectively.

For reasons of computational efficiency, the generalized variance is the most common optimality criterion considered and the designs used in this paper are D-optimal.

One typically compares two designs by looking at the ratio of their D-optimal values. We say that the efficiency of design 1 relative to design 2 is given by

$$\left(\frac{\det(C_1)}{\det(C_2)}\right)^{1/p} \tag{1}$$

where p is the number of parameters to be estimated in a CLM. Some authors (for example, Huber and Zwerina,1996; Sandor and Wedel, 2002) give the D-error of a design, which they define to be $(\det(C^{-1}))^{(1/p)}$, and they call the inverse of D-error "D-efficiency". Unfortunately, however, this measure of D-efficiency does not directly compare two designs. Efficient designs give more accurate parameter estimates from the same number of respondents compared to less efficient designs.

If the optimal design is known in a given context, we can compare any proposed design to this optimal design. Since no design can be better than the optimal design (by definition) the efficiency relative to the optimal design is just referred to as the efficiency of the design. Designs which have efficiencies over 0.9, say, relative to the optimal design are sometimes said to be near optimal. Typically one also wants to estimate main effects independently of each other, which requires a C^{-1} matrix that is diagonal, or at least block diagonal, where these blocks correspond to the main effects, so that different main effects are estimated independently of each other. For example, one can obtain such a C^{-1} matrix by using an orthogonal main effects plan (hereafter, "OMEP", or a "resolution 3 design") as a starting design, and then adding appropriate generators to the starting design to create the choice sets. Burgess and Street (2005) and Street and Burgess (2007) formally discuss the properties of "generator developed" designs, and derive several key results associated with maximizing the determinant of C under the null hypothesis. These efficient designs often have a small number of choice sets, making them ideal candidates for designing DCEs that can be used to model the choices of individual decision makers. Although not globally optimal, the generator developed designs are optimal in a region around the origin (Street and Burgess, 2007), and can estimate all main effects in any part of the β space. This makes them a good choice for the initial investigation in virtually all generic choice problems.

A CLM for a generic DCE is specified in terms of differences in attribute levels (see, e.g., Louviere et al., 2000, pp 112-115), which can be seen by linearizing the MNL model:

$$P(i \mid A) = exp(V_i)/\sum_j exp(V_j), \tag{2}$$

where $P(i \mid A)$ is the probability of choosing the i-th option from choice set A, and V_i, V_j are systematic utility components of utility for choice options i and j. Typically, V_i and V_j are specified as linear-in-the-parameters utility expressions, and for generic utility expressions, each systematic utility parameter is assumed constant across all options.

Thus, as shown below, we can express the systematic utilities of options i and j as linear in the parameters forms, where the β_0 represent generic intercept effects for choice options i and j, and the β_k represent preference parameters associated with the k-th attribute, or X_k's.

$$V_i = \beta_0 + \sum_k \beta_k X_{ki}, \text{ and } V_j = \beta_0 + \sum_k \beta_k X_{kj}. \tag{3}$$

If we let option m be a reference option, and we consider the odds ratio of choosing option i relative to option m, we obtain

$$P(i \mid A)/P(m \mid A) = [exp(V_i)/\sum_j exp(V_m)]/[exp(V_j)/\sum_j exp(V_m)]. \tag{4}$$

If we take the natural log of equation (4), we obtain:

$$log_e[(P(i \mid A)/P(m \mid A)] = \sum_k \beta_k (X_{ki} - X_{km}), \tag{5}$$

where log_e is the natural logarithm. If utility functions are not generic, one or more terms should be subscripted to allow parameters to differ for some options (e.g., commuter rail and bus may exhibit different sensitivities to travel time).

Equation (3) with equation (5) shows that a CLM is a difference-in-attributes model. The differences are expressed relative to a reference or base option, which in this case is m. If the attributes are categorical, rather than quantitative, the differences represent contrasts between levels. The designs developed by Burgess and Street (2003, 2005) and Street and Burgess (2004a,b, 2007) allow one to construct optimal or nearly optimal designs for generic DCEs, which in turn allow one to estimate generic CLMs of the general form shown in equation (5).

3 Maximizing the Amount of Choice Information in Each Choice Set

Louviere et al. (1999) and Hensher et al. (1999) discuss various ways to collect responses from persons consistent with RUT, showing how various response modes can be used to obtain extra choice information that can be expanded into implied choices (see also, Brazell et al., 2006). For example, if there are J options in a particular choice set, and a person reports their most and least preferred options in that set, this can be expanded to several implied choice sets. It is instructive to illustrate this process and its connection with choice models.

Consider a simple case of four hot drinks described by two attributes, each with two levels: type of drink (coffee, tea) and milk (yes, no). All possible combinations of the four drinks are a) coffee with milk, b) coffee black, c) tea with milk, and d) tea black. Following Louviere and Woodworth (1983), we use a 2^J (J=1, . . . , 4) design to put the four drinks into choice sets ("2" = if a drink is "present" or "absent" in each set). The 16 possible sets are in Table 1. The first line of the table shows a hypothetical preference ranking of someone asked to make choices in ALL subsets (except the empty set). The bottom two lines in the table shows the summed choices of each option that one should obtain if that person chooses perfectly consistently with their rank order.

As Louviere and Woodworth (1983) note, the totals at the bottom of Table 1 contain ALL the information in the choice data. One can view these totals in two ways: a) as marginal choice frequencies associated with each choice option defined by each combination of drink and milk; that is, each total provides information to estimate an alternative-specific constant for each option; b) as information needed to decompose the choices into a generic indirect utility function represented by main effects for drink type and milk (and, potentially, a drink type by milk interaction).

One also can disaggregate choices set-by-set and estimate model parameters from disaggregate choices. Regardless of whether one estimates parameters from summed choice counts or disaggregated choice indicators (1, 0), parameter estimates will be the *same* up to scale. That is, the estimates will be perfectly pro-

Table 1: All Possible Choice Sets For Four Designed Drink Options

Preference order →	1	2	3	4
Choice set	**Black coffee**	**White coffee**	**Black tea**	**White tea**
1 (null or empty set)	0	0	0	0
2	0	0	0	1
3	0	0	1	0
4	0	0	1	1
5	0	1	0	0
6	0	1	0	1
7	0	1	1	0
8	0	1	1	1
9	1	0	0	0
10	1	0	0	1
11	1	0	1	0
12	1	0	1	1
13	1	1	0	0
14	1	1	0	1
15	1	1	1	0
16	1	1	1	1
Expected totals, non-null sets	8	4	2	1
Expected totals, non-null, non-single sets	7	3	1	0

portional, with the constant of proportionality reflecting the degree of variability in the disaggregated choices relative to the aggregated choices (e.g., Louviere et al., 1999; Swait and Louviere, 1993).

To illustrate this relationship, we estimate the main effects of the CLM using a weighted least squares (WLS) regression approach to fit the aggregate choices; this estimation approach produces consistent, but inefficient, estimates (Louviere and Woodworth, 1983). It is well-known that WLS is the first-step estimator in the iteratively reweighted least squares (IRLS) approach to obtaining the maximum likelihood estimates (e.g., Green, 1984); and comparisons of maximum likelihood estimates and WLS estimates for CLMs typically suggest minor differences. Hence, we use WLS to estimate the parameters of CLMs at several points in this paper because WLS is easy to do and most analysts should have access to suitable estimation software. Results are in Table 2. Naturally, because the underlying process is deterministic, the WLS results explain 100% of the choice totals.

Now, consider a larger number of drink combinations described by type of drink (coffee, tea, coke, water), price ($1.00, $1.50, $2.00, $2.50) and container (bottle, can). We construct 16 choice options using an orthogonal main effects plan (OMEP) to make drink descriptions, as in Table 3. Instead of assigning the

Table 2: A WLS Regression Results

Effect	**Estimate**	**StdErr**	**t**	**P(t)**
Constant	1.040	.000	.	.
Drink	-.693	.000	.	.
Milk	-.347	.000	.	.

Table 3: Design to Construct Options

Option	**X1**	**X2**	**X3**
1	0	0	0
2	0	1	0
3	0	2	1
4	0	3	1
5	1	0	0
6	1	1	0
7	1	2	1
8	1	3	1
9	2	0	1
10	2	1	1
11	2	2	0
12	2	3	0
13	3	0	1
14	3	1	1
15	3	2	0
16	3	3	0

16 drink descriptions to choice sets of varying size using a 2^J design as in Table 1 (here J=16), we assign them to sets of fixed size. One way to do this is to use a balanced incomplete block design (BIBD) to put the 16 drink options into sets (see, e.g., Street and Street, 1987).

That is, we can use a BIBD to assign the 16 attribute combinations in Table 3 to 20 choice sets of size four, as shown in Table 4. This produces a DCE where each choice option occurs five times and co-occurs with every other option once. Now, suppose we let the utility of each level of each attribute be as follows:

- Drink Type (coffee=1.25, tea=-2.5, coke=0.0, water=0.0);
- Price ($1.00=2.50, $1.50=1.00, $2.00=0.25, $2.50=-0.25); and
- Container (bottle=1.25, can=-2.50).

We next calculate the (deterministic) expected utility of each choice option from the above utility values in each of the 20 choice sets, and use this information to rank the options in ways discussed below.

Table 4: Using A BIBD to Construct 20 Choice Sets for 16 Designed Options

Choice Set	Options per set			
1	2	5	8	14
2	1	5	6	7
3	5	9	12	16
4	4	5	11	15
5	3	5	10	13
6	1	2	3	4
7	2	6	9	11
8	2	7	13	16
9	2	10	12	15
10	1	8	9	10
11	6	8	13	15
12	4	7	8	12
13	3	8	11	16
14	1	14	15	16
15	3	6	12	14
16	7	10	11	14
17	4	9	13	14
18	1	11	12	13
19	4	6	10	16
20	3	7	9	15

First, we rank the options in each choice set and estimate an ordinal regression model from the rankings, as shown in Table 5. These estimates correlate highly with the "true" utilities within each attribute. Next we use only the most and least preferred (top and bottom ranked) data, and estimate another ordinal regression model (i.e., we allow an incomplete ranking). These model estimates correlate highly with the estimates from the full rank ordinal regression and the "true" utility values for each attribute level. Finally, we explode the data using the Luce and Suppes (1965) approach (Chapman and Staelin, 1984), which also yields estimates that correlate highly with the "true" estimates. For completeness, we also estimate an exploded logit model using only the information about most and least preferred (bottom and top ranks) using the marginal likelihood method to handle ties; this also yields similar model estimates. We omit both exploded logit results in the interests of space.

Now, consider what happens if we "pretend" that we have more data than only the ranks, but do this in a very systematic and structured way. In particular, we "pretend" that because we know the preference order in each of the 20 choice sets in Table 4, we can apply that order to predict the expected choices in Table 1. That is, in each of the 20 choice sets, if an individual chooses consistently with their preference ranking, we expect the choice counts for options to be as follows (in parentheses): rank 1(8), rank 2(4), rank 3(2), rank 4(1). Hence,

Table 5: Model Estimates from Ordinal Regression analysis of Ranks

	Effect	**Estimate**	**StdErr**	**Wald**	**P(Wald).**
Threshold	rank2 = 1	0.6214	0.7085	0.7691	0.3805
	rank2 = 2	2.7376	0.7803	12.3104	0.0005
	rank2 = 3	5.4173	0.9704	31.1635	0.0000
Location	x1=1	1.4157	0.7069	4.0100	0.0452
	x1=2	-2.5177	0.7672	10.7699	0.0010
	x1=3	0.3723	0.6421	0.3362	0.5620
	x1=4	0.0000	.	.	.
	x2=1	2.5164	0.8273	9.2530	0.0024
	x2=2	0.6841	0.6898	0.9834	0.3214
	x2=3	0.5096	0.6648	0.5877	0.4433
	x2=4	0.0000	.	.	.
	x3=1	4.6075	0.7284	40.0161	0.0000
	x3=2	0.0000	.	.	.

Table 6: WLS Model Estimated from Simulated Data

Effect	Estimate	StdErr	T	P(t)
Constant	1.1783	0.1202	9.8051	0.0000
x1fx1	0.1272	0.1221	1.0415	0.3011
x1fx2	-0.5358	0.1557	-3.4421	0.0010
x1fx3	0.1993	0.1271	1.5676	0.1214
x1fx4	0.2094			
x2fx1	0.3587	0.1248	2.8746	0.0053
x2fx2	0.0739	0.1423	0.5191	0.6053
x2fx3	0.0171	0.1400	0.1221	0.9032
x2fx4	-0.4496			
x3fx1	0.4905			
x3fx2	-0.4905	0.0512	-9.5717	0.0000

one can simply use the ranking of the four options in each set to estimate the expected totals in the design in Table 1 *for each choice set* in Table 4. Now we use the expected choice totals in each choice set in Table 4 to estimate a CLM model from all 20 choice sets, as shown in Table 6. We again estimate the parameters of the implied indirect utility function of the CLM using WLS (Louviere and Woodworth, 1983). If we use these model estimates to predict the "true" utilities, the correlation between the "true" utilities and the predicted utilities is 0.955, with little evidence of any systematic deviation from a line of perfect fit. Thus, the model estimates are highly linearly related to the calculated utilities, and each estimate in Table 6 is highly correlated with its "true" utility (Table 4).

The estimates in Table 6 correlate highly with the ordinal regression estimates

(>0.99) in Table 5, with a slope of approximately 0.25, and a zero intercept. The WLS estimates exhibit more variability than the ordinal regression estimates, but the parameters are approximately the same up to scale. Unlike ordinal regression estimates, the WLS CLM estimates allow one to predict choice probabilities directly. Moreover, the estimated utilities correlate highly with the "true" calculated utility scores (=0.955), and a graph of observed vs predicted shows little systematic variation from a perfect fit.

Our empirical results to this point reflect the fact that the estimation results are robust to several logical ways of expanding ranking data and estimating models from the expanded data. Thus, one can estimate models for individuals using several estimation methods. Our interest in this paper is ***not*** on making ***inferences*** about model effects for a particular person, but instead on obtaining consistent model estimates for that person. Of course, one might well want to estimate a model for a particular "important" person, in which case, inference would be an issue; of course, in that case one might be motivated to estimate the model parameters using maximum likelihood. The key takeaway here is that one does not need to do this to obtain consistent estimates.

Now we focus on estimating CLMs using WLS for the case where one "pretends" that a full ranking of objects in each choice set gives the expected number of choices in all possible subsets of that choice set. Choice set sizes of three to five often are used in DCEs; hence, if one adds extra questions about most and least preferred options from subsets of remaining options in each set, one can obtain a full ranking of the options. That is, as previously illustrated by the examples in Tables 1 and 4, if an individual is perfectly consistent, and must choose an option in all possible subsets of choice sets, a ranking of J options allows one to infer the implied most preferred option in every non-empty subset (there are $2^J - 1$ such sets). To summarize, several ways can be used to "expand" choice data based on full or partial rankings in order to estimate models that are consistent with random utility theory. Several "expansions" yield estimates that are highly related to the "true" estimates in the cases discussed so far.

To this point, the data in the examples were errorless. Thus, it is worth noting that "randomness" in choices for real individuals will come from the fact that they may make mistakes in reporting their ranking in each choice set *a la* dePalma et al. (1994), which leads to non-systematic errors within and between choice sets. So, a particularly simple way to estimate models for individuals is to treat the implied number of choices for each option as weights, and estimate a weighted CLM model for each individual. One also can estimate a CLM for each individual using WLS as previously illustrated (details of data set up are in the **Empirical Illustration** in the next section).

Table 7: A Set of Optimally Efficient Triples Using Street-Burgess Design Theory

	Option a			**Option b**			**Option c**		
Choice Set	**Price**	**Meal**	**Drink**	**Price**	**Meal**	**Drink**	**Price**	**Meal**	**Drink**
1	0	0	0	0	1	1	1	0	1
2	0	1	1	0	0	0	1	1	0
3	1	0	1	1	1	0	0	0	0
4	1	1	0	1	0	1	0	1	1

Table 8: Utility, Rank and Weight Calculations Based on True Utility Function

Triples	**Price**	**Meals**	**Drinks**	**Utility**	**Rank**	**Weights**
1a	7	-1	-1	-1.45	3	1
1b	7	1	1	3.05	1	4
1c	9	-1	1	-0.15	2	2
2a	7	1	1	3.05	1	4
2b	7	-1	-1	-1.45	3	1
2c	9	1	-1	1.35	2	2
3a	9	-1	1	-0.15	2	2
3b	9	1	-1	1.35	1	4
3c	7	-1	-1	-1.45	3	1
4a	9	1	-1	1.35	2	2
4b	9	-1	1	-0.15	3	1
4c	7	1	1	3.05	1	4

4 The Case of Optimally Efficient DCEs

Before proceeding to the **Empirical Illustration**, we use two examples to show how the prior discussion applies to the case of optimally efficient OMEPs based on Street-Burgess (SB) design theory (Street and Burgess, 2007). The first hypothetical DCE involves lunch choices, where each lunch consists of a meal, a drink and an associated price. Lunch attributes levels are meal (pizza slice or chicken burger), drink (water or Coca-Cola) and prices ($7 or $9). An optimally efficient set of triples for this $2 \times 2 \times 2$ is shown in Table 7, constructed using the ideas in Burgess and Street (2003).

Now, we calculate a (deterministic) utility for each option using a "true" indirect utility function: U = 1.5 - 0.1(price) + 1.5(meal) + 0.75(drink). The recoded design matrix (0,1 in Table 7 recoded to -1, 1 in Table 8) and the calculated utilities associated with each option in each set are shown in Table 8.

In this example, the rank order preferences are as given in Table 9.

We expand the choices in each set into 7 implied choice (sub)sets with the logic in Table 1 and related discussion (there are 7 implied sets for 3 options per set), which yields the "weights" in the last column of Table 8. As before, we use WLS to estimate an implied CLM model, with results in Table 10. This is a very

Table 9: Rank order preferences from first example

	a	b	c
Choice Set 1	3	1	2
Choice Set 2	1	3	2
Choice Set 3	2	1	3
Choice Set 4	2	3	1

Table 10: WLS Estimation Results for CLM Model Analysis on Data in Table 8

Effect	**Estimate**	**StdErr**	**T-Stat**	**P(T)**
Price	-0.00015	0.22778	-0.00065	0.99948
Meal	0.52192	0.22553	2.31422	0.02066
Drink	0.26572	0.22725	1.16932	0.24227

Pizza=-1; chicken=1; water=-1; Coke=1
$\rho^2 = 0.50$

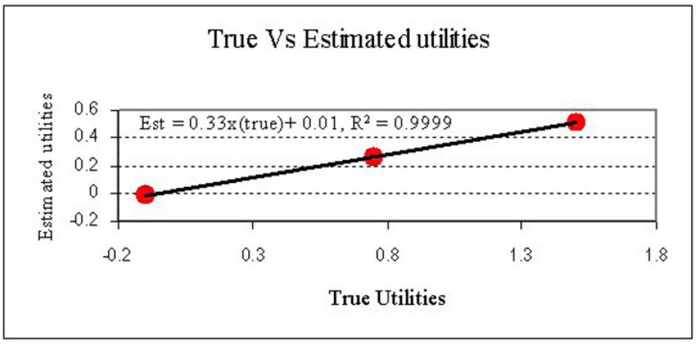


Figure 1: True Versus Estimated Utilities From Weighted CLM Analysis

small experiment, with a very small sample size, so it is not surprising that only meal is significant.

The relationship between true and estimated utilities is in Figure 1. We also estimated a) an exploded logit model (Chapman and Staelin, 1982), and b) a conditional logit model using the ranks to create all pairs of choices in each set (Horsky and Rao, 1984). Both results give model estimates proportional to the WLS estimates, standard errors are not identified as the models are saturated. We omit the results in the interests of brevity.

The second example involves a larger and more realistic experiment, which is a 4^3 design for cross-country flights, with the attributes being fare ($350, $450, $550, $650), total travel time (4, 5, 6, 7 hours) and number of stops (0, 1, 2, 3). This example is a smaller version of the first empirical study discussed in the next

Table 11: Attribute Levels and "True" Utilities Used in Example 2

fare	**utility**	**stops**	**utility**	**time**	**utility**
350	-4.625	0	-5.0	4	-5.0
450	-5.375	1	-5.5	5	-5.5
550	-6.125	2	-6.0	6	-6.0
650	-6.875	3	-6.5	7	-6.5

section (i.e., there are fewer attributes). We used SB design theory (Burgess and Street, 2003) to produce an optimal set of quadruples for the 4^3, which results in 16 choice sets with four options in each set. This design has the property that all levels of each attribute appear in each choice set. We also use this property in a second illustration to show how the WLS estimation approach can identify lexicographic behavior.

As before, we begin by creating a "true" (deterministic) indirect utility function, and use it to calculate the expected utilities for each option in each choice set. These "true" utilities are shown in Table 11. We next rank each option in each choice set using the calculated utilities, and use WLS to estimate a CLM from the data as described earlier. The model estimates from this give predicted utility scores that correlate highly with the "true" utility scores (>0.94). As before, we estimated ordinal regression and exploded logit models from full and partial rankings; both yield estimates that correlate highly with the "true" estimates. We omit these in the interests of space.

Now we apply the same ideas to a lexicographic example where a hypothetical person ranks the items in order of fare levels, and because each level of fare appears in each choice set, one has a full ranking based only on fare. Not surprisingly, this produces a perfect fit, giving estimates for each level of fare, with exactly zero estimates for all remaining attribute levels, as shown in Table 12. In our empirical experience, simple inspection of each individual's WLS estimation results allows one to easily "see" lexicographic choices: there are non-zero estimates for the attribute in question, with all other attribute estimates exactly equal to zero. Our empirical work with real subjects to this point suggests that attributes exhibiting lexicographic effects are almost always brand names and price levels.

We now apply these ideas to two empirical examples to show that the approach works well for estimating models for single individuals in real field applications.

5 Empirical Illustration

5.1 Study 1

The first example is a DCE for choices among different flight options for transcontinental trips like Miami to Seattle, Moscow to Vladivostok or Sydney to

Table 12: Lexicographic Model Estimates

Effect	Estimate	StdErr	t	Sig.
Fare=350	-2.0794	0	.	.
Fare=450	-1.3863	0	.	.
Fare=550	-0.6931	0	.	.
Fare=650	0.0000	.	.	.
stops=0	0.0000	0	.	.
stops=1	0.0000	0	.	.
stops=2	0.0000	0	.	.
stops=3	0.0000	.	.	.
Time=4	0.0000	0	.	.
Time=5	0.0000	0	.	.
Time=6	0.0000	0	.	.
Time=7	0.0000	.	.	.

Table 13: Attributes and Levels Used in DCE for Flights

Attribute	Levels
Round-Trip Air Fare	**\$350, \$450, \$550, \$650**
Number of Stops	**0, 1, 2, 3**
Total Travel Time	**4, 5, 6, 7 hours**
Type of Airplane	**B-737, B-717**
In-Flight food & Beverage	**Beverage only, Beverage + Hotmeal**
Airline	**Southwest, Northwest**

Broome, which is a larger version of the experiment immediately above (i.e., more attributes). Participants were 12 student volunteers who completed the DCE as part of a choice modeling class exercise used to illustrate the design and analysis of DCEs. Each flight is described by three 2-level attributes and three 4-level attributes (See Table 13):

We used the SB design approach to create 16 choice sets, each with four generic choice options. Task instructions and example choice set are in Appendix A, which shows that each individual answered three questions about each choice set plus a "not fly" question (***we do not analyze the data from the "not fly" question in this paper***).

Appendix B shows the data in "stacked" format with each choice option in each choice set represented as a data record. Weights used to estimate CLM models are calculated from the rank order of the choice options in each choice set, and shown in Appendix B. As discussed earlier, weights are calculated for each rank for $J = 4$ as follows: rank 1 = 8, rank 2 = 4, rank 3 = 2 and rank 4 = 1. The dependent variable is the column labeled "alt". In this case, we use Salford Systems LOGIT software to estimate weighted CLM models for each person using maximum likelihood, but any CLM estimation software that allows weighting of choice responses can be used, or one can use the WLS approach described earlier. Table 14 contains estimation results for the 12 people. The indirect utility expressions were specified in a general way using effects codes to estimate parameters for each attribute level; thus, these results show that the approach can estimate non-linear attribute effects. Significant attribute level effects are denoted as follows: a) effects exhibiting t-statistics with probability $<$ 0.01 under the null are marked in ***bold italics***, b) effects exhibiting t-statistics with probability $0.05 < t < 0.01$ are marked in **bold**, and c) effects exhibiting t-stats with probability $0.10 < t < 0.05$ are in *italics*.

The Table 14 results indicate that each person considered two to three of the six attributes when making their choices. Differences between individuals can be seen in Figure 2, which graphically displays the Table 14 results. The graphs show that individuals reacted by and large as expected to fare, number of stops and travel time. Individuals differ greatly in preferences for food, type of airplane and airline. However, the estimated utilities for fare, number of stops and time have a range close to 2.0, while estimated utilities for type of plane, food and airline have a range of approximately 0.4. So, on average, numerical attributes had almost five times as much effect on preferences as the qualitative attributes, but the qualitative attributes exhibited much more preference heterogeneity, as might be expected.

Naturally, our results beg the question of whether they are meaningful and valid. We test for empirical meaningfulness by undertaking a cross-validity test that involves predicting the observed first choices (i.e., the answers to question 1 in each choice set; these are the data typically used to estimate models from DCEs) of each person in each choice set using their estimated parameter values to calculate the utility values associated with each option in each choice set. Each of the 12 persons evaluated 16 choices sets, so there are 192 traditional “first choice” observations available, treating all persons and first preference choices as a single dataset. We use each person’s estimated parameters in Table 14 to predict the expected utility score associated with each choice option in each choice set, or a total of 64 predicted utility scores for each person (of which, 192 are the “first choices”). We then use these predicted utility scores as a single explanatory variable to determine how well they fit the first choice data compared with other model options. Note that the predicted utility scores are not linearly dependent across the 12 individuals because the linear expressions differ for each person.

We begin our evaluation of how well the proposed individual model approach works by estimating a “one size fits all” aggregate CLM model from the first choice data. This model has the same specification as the models in Table 14. The initial LL for this model is 266.2, and the final LL at convergence is 109.2. Thus, a simple CLM model fits the data well, with a pseudo r-square of approximately 0.57, which is not surprising in light of Figure 2 that suggest a fair degree of between-person consistency, especially with respect to numerical attributes. We next fit a CLM with a single parameter associated with the predicted individual-level utility scores, which yields a LL at convergence of 85.2, a reduction of 24

Table 14: Estimated Attribute Level Utilities and Significance Levels

Effect	p01	p02	p03	p04	p05	P06	p07	p08	p09	p10	p11	p12
$350	***0.601***	***0.716***	***0.633***	**0.260**	***0.492***	***0.733***	***0.506***	0.089	***0.412***	***0.488***	***0.482***	**0.227**
$450	***0.451***	***0.366***	***0.537***	***0.327***	0.072	***0.371***	***0.456***	0.128	*0.193*	0.103	***0.567***	0.159
$440	*-0.267*	-0.170	***-0.375***	-0.069	0.026	**-0.335**	-0.101	-0.003	-0.113	-0.126	-0.189	-0.165
0 Stop	0.138	-0.141	**0.250**	0.124	0.114	0.049	0.159	***0.477***	0.008	0.137	**0.235**	***0.755***
1 Stop	0.158	-0.154	***0.440***	0.132	***-0.376***	-0.061	0.028	0.177	0.161	***0.616***	0.138	***0.670***
2 Stop	0.028	0.164	**-0.326**	-0.012	0.133	0.070	-0.084	-0.014	0.019	-0.161	0.004	***-0.405***
4hrs	***0.301***	***0.495***	0.112	***0.748***	***0.436***	*0.206*	***0.533***	***0.557***	***0.575***	***0.464***	***0.430***	0.106
5hrs	0.106	***0.341***	**0.247**	***0.468***	***0.527***	**0.269**	***0.339***	***0.604***	**0.278**	0.123	*0.195*	0.064
6hrs	*0.212*	-0.184	-0.168	-0.132	-0.196	-0.094	***-0.392***	-0.160	***-0.393***	-0.166	-0.023	-0.116
Plane	-0.035	0.017	-0.085	-0.005	-0.081	*0.123*	0.011	-0.024	0.039	0.072	0.030	0.060
Food	0.012	-0.013	0.068	0.001	-0.069	*0.121*	0.076	0.002	-0.008	0.050	0.013	-0.052
Airline	*0.117*	-0.095	-0.008	0.018	0.077	-0.045	0.022	-0.010	0.075	-0.055	0.042	0.045

= ***t<.01***; = **.01 < t < .05**; = *.05 < t < .10*

Table 15: Results of CLM Analysis Including Predicted Utility and Attribute Levels

Effect	Estimate	StdfErr	t	P(t)
Fare=$350	0.405	0.387	1.047	0.295
Fare=$450	0.926	0.271	3.416	0.001
Fare=$550	-0.042	0.352	-0.119	0.905
Stops=0	-0.166	0.368	-0.452	0.652
Stops=1	0.685	0.255	2.683	0.007
Stops=2	0.085	0.327	0.259	0.795
Time=4hrs	0.313	0.360	0.870	0.385
Time=5hrs	0.782	0.268	2.917	0.004
Time=6hrs	-0.190	0.312	-0.611	0.541
PlaneType	-0.124	0.174	-0.713	0.476
Food	-0.121	0.154	-0.783	0.434
Airline	-0.327	0.169	-1.943	0.052
Predicted Utility	4.810	0.748	6.430	0.000

LL points. Each individual model has 12 estimated parameters; hence a total of 144 parameters were estimated to predict the vector of utility scores. Thus, the individual-level utility values do not significantly improve the aggregate CLM, probably due to a) high between-person agreement in responses to the numerical attributes, and b) the numerical attributes drive most of the choices (See Table 14 and individual-level graphs in Figure 2).

We next estimated a model that included the predicted individual-level utility scores plus the design attributes specified as in Table 14, namely in effects-coded form. These results are in Table 15, and reveal several highly significant effects associated with some attribute levels.

The results of this analysis are graphed in Figure 3. To our knowledge this is the first time a model of this type has been estimated, so the results warrant further discussion. Essentially, the attribute effects in Table 15 represent the residual choice variation controlling for the mean (predicted) utility of each person. Thus, they estimate the effects of design attributes on unobserved variability after controlling for mean utility. Louviere et al. (2000), Louviere (2001, 2004a,b) and Louviere and Eagle (2006) discuss these types of effects with respect to types of variance components one should expect to find in DCEs. In particular, the cited sources suggest that for numerical attributes like price, time and number of stops, one should expect inverse-U shaped relationships between residual choice variability and attribute levels because individuals will tend to agree more on extreme levels, and disagree more about interior levels, all else equal. Thus, taken as a group, individuals should be more consistent in choices for the extreme (lowest and highest) numerical levels, with least consistency in choice associated with interior levels. Figure 3 shows that this is indeed the case.

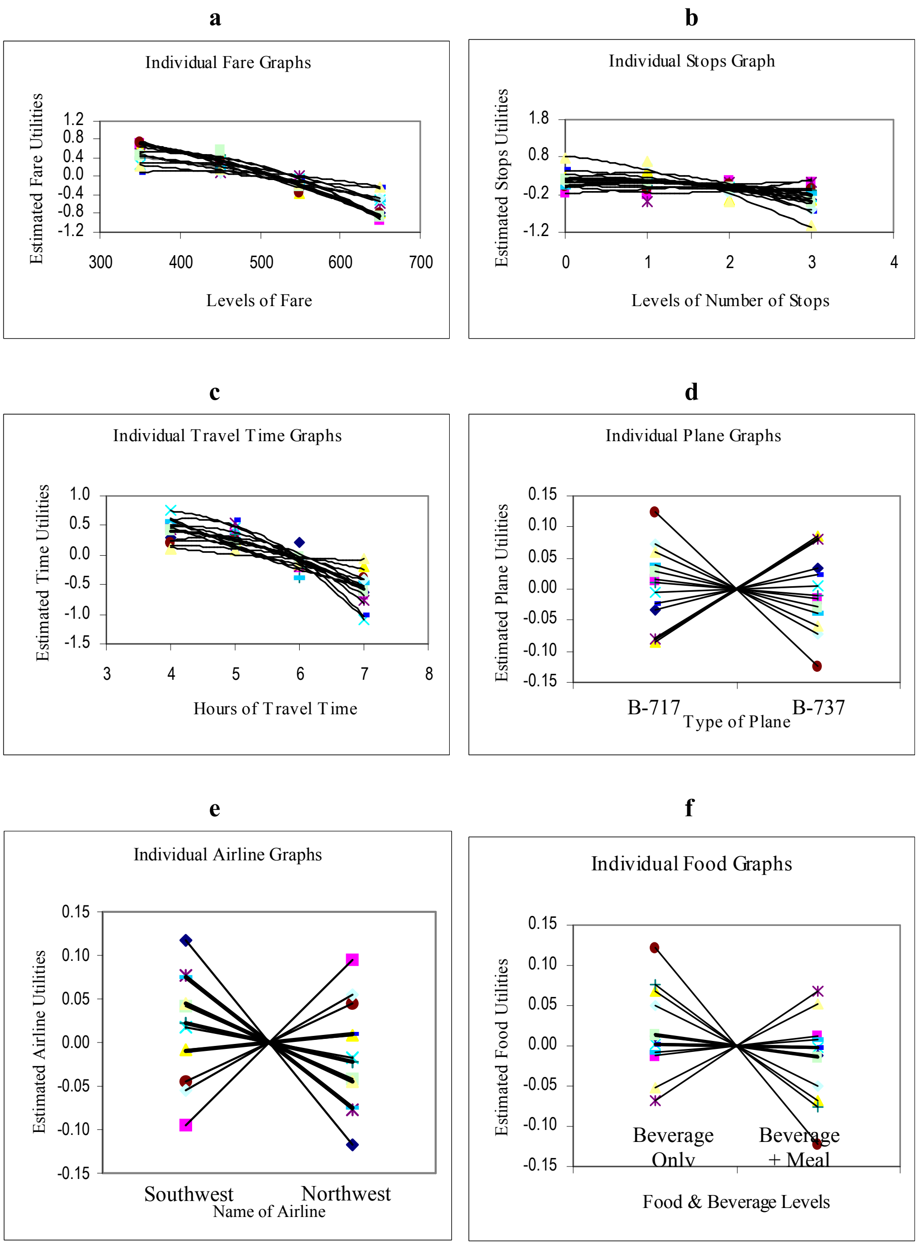


Figure 2: Graphs of Individual Effects from Table 3

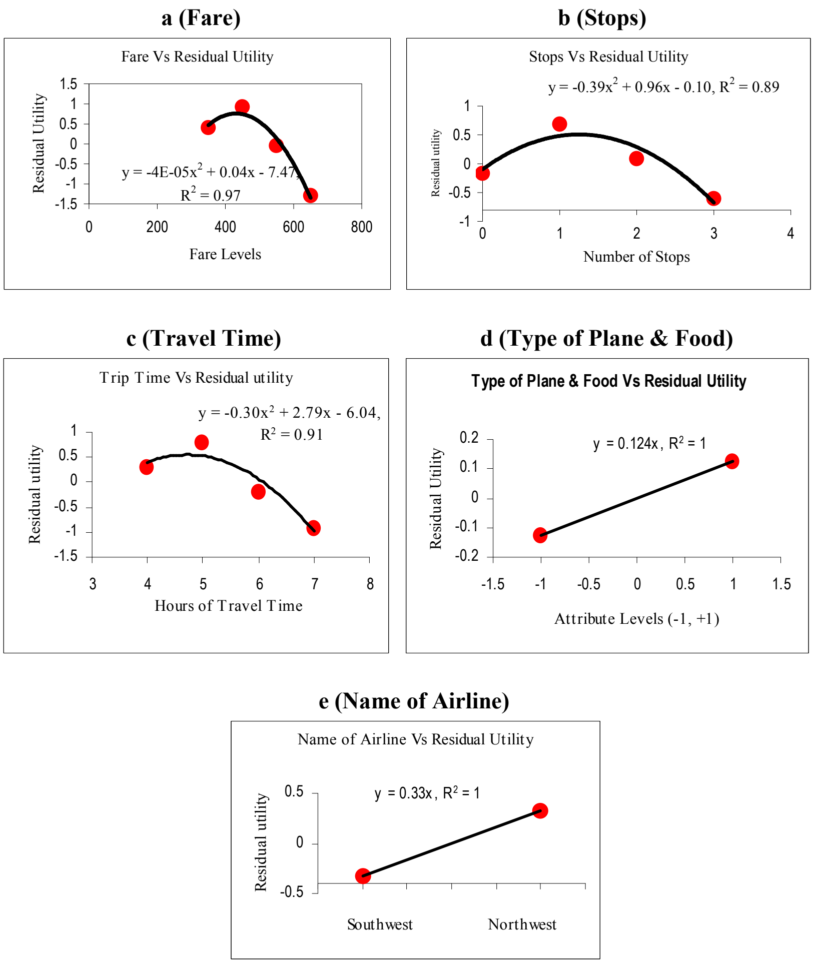


Figure 3: Attribute Effects on Residual Utility

5.2 Study 2

The second empirical study involves two product categories, namely delivered pizzas and packaged juice products. We systematically varied inclusion/exclusion of seven attributes in both categories. Each category had two core attributes always present (brand name and price). Attributes and levels are in Tables 16 and 17, respectively. Attributes/levels for pizzas were based on reviews of local pizza supplier menus and ingredient lists; juice attributes/levels were chosen after examining product descriptions on packages in local supermarkets. Attributes and levels were further refined and reduced using focus groups of individuals from the study population.

Table 16: Attributes/Levels for Delivered Pizza Products

Attribute/Level	0	1	2	3
Brand*	Pizza Hut	Domino's	Eagle Boys	Pizza Haven
Price*	$12	$14	$16	$18
Delivery Time (mins.)	10	20	30	40
♯ of Toppings	1	3		
Free Delivery	No	Yes		
Salad	No	Yes		
Free Drinks	No	Yes		
Free Dessert	No	Yes		
Crust Type	Regular	Thin		

* always present

Table 17: Attributes/Levels for Fruit Juice Products

Attribute/Level	0	1	2	3
Brand*	Berri	Just Juice	Daily Juice	Spring Valley
Price (250 ml)*	$1.00	$1.30	$1.60	$1.90
% real juice	10%	40%	70%	100%
Made From	Concentrate	Fresh		
Vitamin C	Not added	Added		
Sugar	Unsweetened	Sweetened		
Calcium	Not Added	Added		
Package	Glass	Plastic Bottle		
Pulp	Yes	No		

* always present

We used a BIBD to make 14 different "master" conditions (MCs) to include and exclude non-core attributes. We used the SB design approach to make 16 choice sets for each MC; each choice set had 3 or 5 generic choice options. Task instructions and questions were similar to flights, so are not repeated to save space. As with flights, an option to choose no options was a question in each set (<u>but not modeled here</u>). Presence (P)/Absence (blank) of attributes in different conditions are summarized in Table 18.

We used maximum likelihood to estimate a weighted individual-level CLM for each of 560 persons (14 conditions $\times$ 10 individuals per condition $\times$ 2 product categories $\times$ 2 different choice set sizes (3 or 5)). We then used the estimated models to predict the observed choice proportions of each person. Observed choice proportions are calculated by transforming choice rank weights into proportions for each person. So, for example, in the case of triples, the weights are 4, 2, and 1, implying proportions of $4/7$, $2/7$ and $1/7 = 0.571, 0.286, 0.143$. For quintuples,

Table 18: Master Conditions (presence/absence) Design

Fruit Juices			**%real juice**	**madefrom**	**vitamin C**	**sugar**	**calcium**	**packagetype**	**pulp**
DeliveredPizzas	**brand**	**price**	**deliverytime**	**toppings**	**freedelivery**	**freesalad**	**freedrink**	**free dessert**	**crusttype**
Condition									
1	P[1]	P		P		P		P	
2	P	P	P			P	P		
3	P	P			P	P			P
4	P	P	P	P	P				
5	P	P		P			P		P
6	P	P	P					P	P
7	P	P			P		P	P	
8	P	P	P		P		P		P
9	P	P		P	P			P	P
10	P	P	P	P			P	P	
11	P	P				P	P	P	P
12	P	P	P		P	P		P	
13	P	P		P	P	P	P		
14	P	P	P	P		P			P

[1] P = Attribute is Present in the Condition

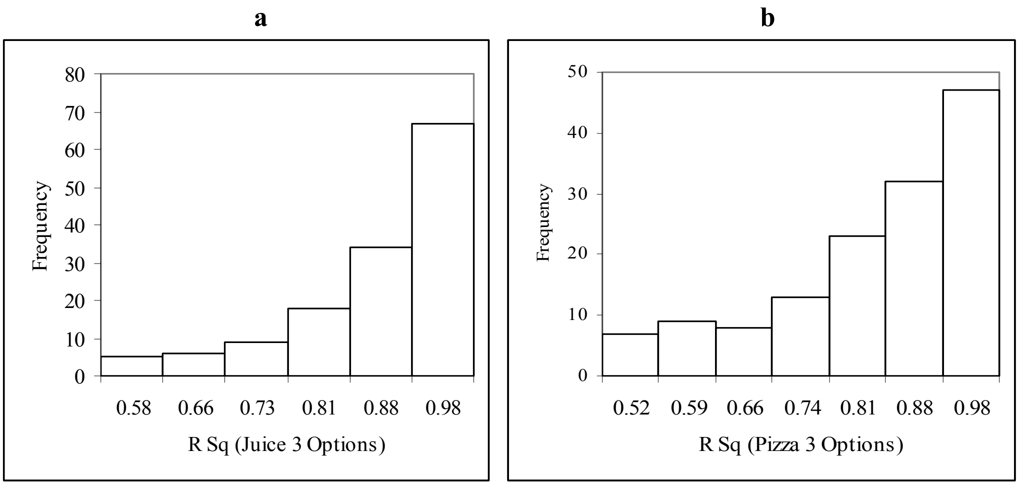


Figure 4: Distribution of R-square measures for pizza and juice conditions

weights are 16, 8, 4, 2 and 1, implying proportions of 16/31, 8/31, 4/31, 2/31 and 1/31 =0.52, 0.26, 0.13, 0,065, 0.032. Three goodness-of-fit measures were calculated based on the observed and predicted choice proportions: a) mean square residual, b) R-square and c) Chi-square. Summaries of the goodness-of-fit measures (i.e., means and medians) are in Table 19; the distribution of R-square measures for pizza and juice conditions is shown in Figure 4

We conducted a further test of how well the individual-level models capture the underlying utilities and choices. Each individual-level CLM was used to calculate a vector of predicted utility values for each person. The null hypothesis in this (sequentially estimated) test is that the design attributes should not be significant once we account for their effects with the predicted utility vector. We estimate two "first choice" CLMs using the disaggregate first choice data from the sample (each MC $\times$ number of choice options $\times$ number of sets $\times$ number of participants). The first CLM model is specified with alternative-specific constants (ASCs) and a single vector of predicted utilities. The second CLM model is specified with ASCs, the predicted utility vector and effects-coded attribute levels. So, this test is similar to that reported for Study 1 using data from 12 people, with the difference being that the null is no collectively significant attribute effects after controlling for the individual-level utility vector. The first and second CLM models are nested, so we used likelihood ratio statistics (LRT) to test model differences. These results are summarized in Table 20. As a whole, the results suggest that the individual-level models capture virtually all the variance in the systematic utility components across the samples, suggesting no bias in

Table 19: Goodness-of-Fit of Individual Level Models

		Pizza		**Juice**	
	Measures*	**Median**	**Mean**	**Median**	**Mean**
3 Options	R **–Sq**	0.84	0.79	0.88	0.84
	Mean Sq	0.0052	0.0067	0.0040	0.0052
	Chi Sq	0.82	1.03	0.60	0.80
5 Options	R **–Sq**	0.80	0.71	0.80	0.74
	Mean Sq	0.0065	0.0092	0.0062	0.0080
	Chi Sq	2.61	3.71	2.28	3.12

* Each number is a summary result of 140 observations (14 conditions × 10 individuals).

the estimates because the coded attribute effects are not significant, indicating that their effects are reasonably well approximated by the individual-level model estimates.

We graphically describe the distribution of sample brand and price estimates for all MCs in Figure 5 using histograms. These plots clearly show heterogeneity in individual estimates; more heterogeneity is associated with MCs with five options per set than MCs with 3 options per set. The extra heterogeneity/variability for five options is likely to be at least partially due to more choice variability associated with evaluating and making choices among five options, which should be more difficult cognitively (see also DeShazo and Fermo, 2002; Swait and Adamowicz, 2001).

6 Discussion and Conclusions

To our knowledge, this paper represents the first demonstration of fairly straightforward ways to model the choices of single individuals for problems of typical size in DCE survey applications. Instead of relying on large numbers of choice sets, which generally is impractical outside of controlled laboratory situations, we proposed an approach that combines optimally efficient generator-developed generic DCEs with extra choice questions to obtain as complete a preference ordering of choice options in each set as possible. The empirical examples used to illustrate the approach showed that we can model individual-level choices in nontrivial cases involving three, four or five choice options per choice set, and six to 10 attributes varying over 2 or 4 levels. We were able to estimate individual-level models in all these conditions.

In the interests of brevity, we do not report data and results for several thousand more individuals in several product categories and conditions; these results yield the same conclusions, namely that we can estimate CLMs for single persons in all cases. It is worth noting that all DCEs were conducted online using webpanels, and it may surprise many to learn that completion rates given acceptance

Table 20: Likelihood Ratio Statistics for CLM Models 1 and 2

	Pizza3 Options	**Pizza5 Options**	**Juice3 Options**	**Juice5 Options**		**Chi-Square**
Conditions	**LRT**	**LRT**	**LRT**	**LRT**	**DF =**	**Critical Value at α = 0.05**
1	3.13	4.20	9.73	2.51	9	16.92
2	5.87	7.64	8.10	7.26	11	19.68
3	2.70	3.06	1.20	8.45	9	16.92
4	9.85	6.19	9.73	8.51	11	19.68
5	4.95	4.63	9.63	5.92	9	16.92
6	4.81	3.01	9.57	5.70	11	19.68
7	7.39	8.09	**18.31**	5.19	9	16.92
8	4.80	4.38	6.79	16.73	12	21.03
9	5.34	2.33	5.73	2.35	10	18.31
10	15.22	12.50	16.66	6.54	12	21.03
11	10.68	3.57	5.45	3.34	10	18.31
12	x[1]	2.63	4.09	7.23	12	21.03
13	6.23	3.34	6.91	6.16	10	18.31
14	x[1]	5.79	x[1]	5.58	12	21.03

[1] cases where unrestricted model failed to invert Hessian.

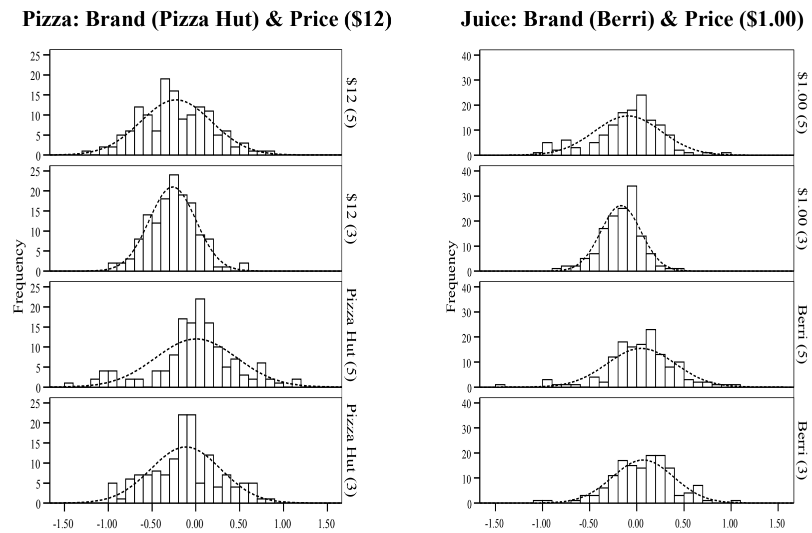


Figure 5: Deviations of Individual Models Coefficients from Aggregate CLM model for Brand and Price (3 = Choice Set Size 3 and 5 = Choice Set Size 5)

of invitation to participate for these surveys were high, varying between 67% and 100%, with a mean of 80%. Analysis of completion rates is the subject of a separate paper that can be obtained from the authors on request. Our empirical experience suggests that designs involving three, four or five choice options are needed to obtain sufficient data to model individuals. Work not reported here using simulated choice data suggests that four is likely an optimum number of options per set, and our empirical experience with this approach also points to four options per set as desirable. Earlier work by Tversky (1964) suggests that three options per set is an optimal number.

Naturally, there are unresolved issues associated with the approach, typical of any new domain of research. For example, we do not know if we can extend the approach to labeled, alternative-specific choice experiments and/or more complex models than CLM. We also do not know the optimum number of choice sets or choice options per set for all problems. However, we can say that we have used the approach on problems as small as six attributes and three options per set with 16 choice sets per person, and as large as 13 attributes and five options per set with 32 choice sets per person. All applications were implemented using online panels; all exhibited high response rates and we were able to estimate models for every individual. We hope that the "proof of concept" illustration as well as the more extensive study of over 500 individuals will stimulate further study of efficient choice designs combined with additional choice questions as a way to tackle the long-standing and important problem of modeling individual choices.

Acknowledgements

We gratefully acknowledge the support of ARC Discovery Grant ♯DP0343632, entitled "Modelling the Choices of Individuals" and Social Sciences and Humanities Research Council (SSHRC) of Canada Grant ♯046288". We also gratefully acknowledge helpful suggestions and comments from Don Anderson, Michael Keane, Joffre Swait and Ken Train on earlier versions of this paper.

A Task Instructions and Example Choice Sets

A.1 Instructions

The following pages contain 16 different scenarios that describe possible flights that are available for you to fly from Philadelphia to Seattle on business. Each flight is described by a round-trip fare, the number of stops en route, the total elapsed travel time to get from the Philadelphia Airport to the Seattle Airport, type of equipment or airplane, type of in-flight food and beverage service and the airline offering the flight.

Please assume that you have a substantial interest in the business that is the reason for your travel, and hence you directly or indirectly will bear the costs of the trip. As you probably are aware, airlines use information about flight choices to decide which flights to offer at what times and for what fares, so your answers will play a role in determine future flight options. The airlines already have learned that you need to leave and arrive at convenient times, and hence you should assume that the options offered to you depart and arrive at times convenient for your needs.

Your task is simple:

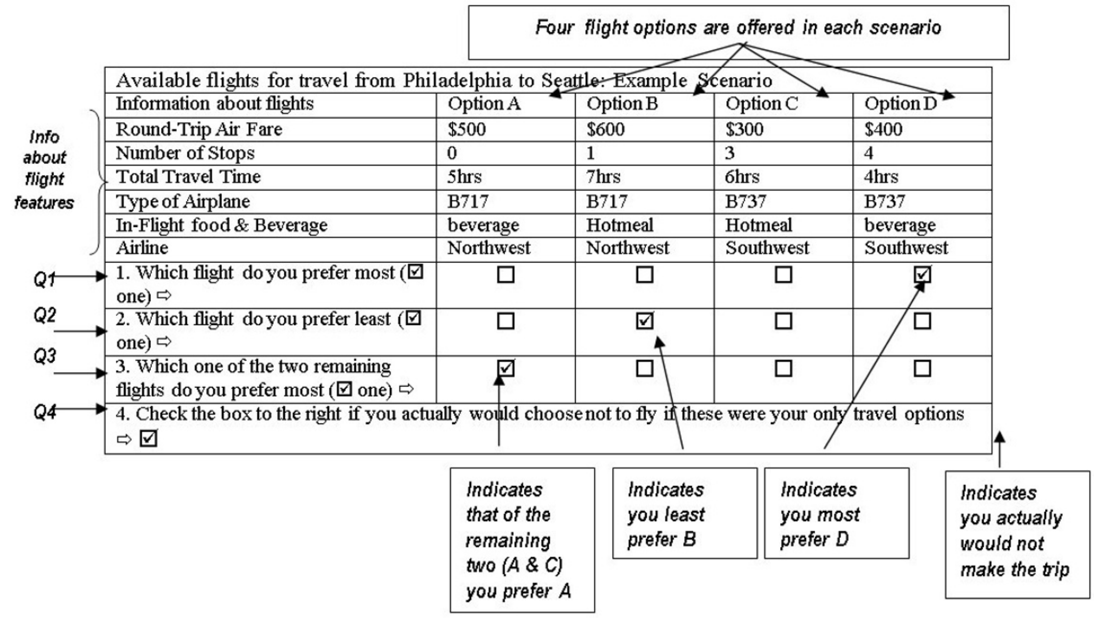
1. Evaluate each of the flight options,
2. Decide which one you prefer the most,
3. Decide which one you prefer the least,
4. Decide which of the remaining two flight options you prefer the most,
5. And decide whether you actually would make the trip at all if the only options that you could choose were the ones offered.

We ask you four questions about each scenario that reflects the task above. Please insure that you answer EVERY question. Each question requires you to check ONLY ONE BOX, and so be sure that in each scenario you have checked AT LEAST THREE BOXES, but DO NOT CHECK more than FOUR BOXES.

A.2 Example choice set

The next page provides an example of how to respond. So, please turn the page and look at the example now.

Example of how to respond to SP task



Available flights for travel from Philadelphia to Seattle: Example Scenario				
Information about flights	Option A	Option B	Option C	Option D
Round-Trip Air Fare	$500	$600	$300	$400
Number of Stops	0	1	3	4
Total Travel Time	5hrs	7hrs	6hrs	4hrs
Type of Airplane	B717	B717	B737	B737
In-Flight food & Beverage	beverage	Hotmeal	Hotmeal	beverage
Airline	Northwest	Northwest	Southwest	Southwest
1. Which flight do you prefer most (☑ one) ⇨	☐	☐	☐	☑
2. Which flight do you prefer least (☑ one) ⇨	☐	☑	☐	☐
3. Which one of the two remaining flights do you prefer most (☑ one) ⇨	☑	☐	☐	☐
4. Check the box to the right if you actually would choose not to fly if these were your only travel options ⇨ ☑				

B Data Layout For Individual Level Analyses

Tables 21 and 22 show the data layout for the individual level analysis, with 16 choice sets, 4 alternatives, and where sxx gives the participant number.

Table 21: Data layout part I

set	**alt**	**s01**	**s02**	**s03**	**s04**	**s05**	**s06**	**s07**	**s08**	**s09**	**s10**	**s11**	**s12**	**fare**	**stops**	**time**	**plane**	**food**	**brand**
1	1	1	4	4	4	4	4	4	4	2	3	4	2	550	1	7	-1	-1	-1
1	2	2	3	3	3	3	3	3	1	3	2	3	1	650	0	5	1	-1	1
1	3	3	2	2	2	2	1	2	3	4	1	2	4	350	2	6	1	1	-1
1	4	4	1	1	1	1	2	1	2	1	4	1	3	450	3	4	-1	1	1
2	1	1	3	4	3	3	4	3	2	3	4	4	4	650	3	4	-1	-1	-1
2	2	4	1	3	2	1	1	4	3	1	3	1	3	350	2	6	1	-1	1
2	3	2	4	2	4	4	2	1	4	4	2	2	1	450	0	7	1	1	-1
2	4	3	2	1	1	2	3	2	1	2	1	3	2	550	1	5	-1	1	1
3	1	4	1	2	2	2	3	2	2	2	2	2	3	350	2	6	-1	-1	-1
3	2	3	2	1	1	3	1	1	1	1	1	1	1	450	1	4	1	-1	1
3	3	2	3	4	3	1	2	3	3	3	3	3	4	550	3	5	1	1	-1
3	4	1	4	3	4	4	4	4	4	4	4	4	2	650	0	7	-1	1	1
4	1	2	4	4	3	3	3	4	2	3	4	3	3	650	2	5	-1	1	1
4	2	1	2	1	4	4	1	2	4	2	1	2	1	350	1	7	1	1	-1
4	3	4	1	3	1	1	2	1	3	1	2	1	4	450	3	4	1	-1	1
4	4	3	3	2	2	2	4	3	1	4	3	4	2	550	0	6	-1	-1	-1
5	1	2	1	2	3	4	1	3	3	4	4	4	4	450	3	6	1	1	-1
5	2	4	2	4	1	1	2	1	2	1	2	1	3	550	2	4	-1	1	1
5	3	1	4	3	2	2	4	4	1	2	3	3	2	650	0	5	-1	-1	-1
5	4	3	3	1	4	3	3	2	4	3	1	2	1	350	1	7	1	-1	1
6	1	2	4	3	3	3	4	3	3	3	3	4	2	650	1	6	1	-1	1
6	2	4	1	1	1	1	1	1	1	1	1	1	1	350	0	4	-1	-1	-1
6	3	3	2	2	2	2	2	2	2	4	2	2	3	450	2	5	-1	1	1
6	4	1	3	4	4	4	3	4	4	2	4	3	4	550	3	7	1	1	-1
7	1	3	1	1	1	1	2	1	1	1	3	1	1	450	0	5	-1	-1	-1
7	2	1	4	4	4	4	3	4	4	4	4	4	4	550	3	7	1	-1	1
7	3	2	3	3	2	3	4	3	2	3	1	3	2	650	1	4	1	1	-1
7	4	4	2	2	3	2	1	2	3	2	2	2	3	350	2	6	-1	1	1
8	1	2	2	3	3	2	3	2	3	3	3	3	4	550	3	5	1	-1	1
8	2	1	4	4	4	4	4	4	4	4	4	4	3	650	2	7	-1	-1	-1
8	3	4	1	1	1	1	1	1	1	1	1	1	1	350	0	4	-1	1	1
8	4	3	3	2	2	3	2	3	2	2	2	2	2	450	1	6	1	1	-1

Table 22: Data layout part II

set	alt	s01	s02	s03	s04	s05	s06	s07	s08	s09	s10	s11	s12	fare	stops	time	plane	food	brand
9	1	1	4	3	4	4	2	4	4	4	4	4	2	650	0	7	1	1	-1
9	2	3	1	2	2	1	1	1	2	1	2	2	4	350	3	5	-1	1	1
9	3	4	3	1	3	3	3	3	3	3	1	1	1	450	1	6	-1	-1	-1
9	4	2	2	4	1	2	4	2	1	2	3	3	3	550	2	4	1	-1	1
10	1	2	3	3	4	4	3	3	4	3	3	3	3	450	2	7	1	-1	1
10	2	3	2	2	2	2	2	2	2	2	2	2	2	550	1	5	-1	-1	-1
10	3	1	4	4	3	3	4	4	3	4	4	4	4	650	3	6	-1	1	1
10	4	4	1	1	1	1	1	1	1	1	1	1	1	350	0	4	1	1	-1
11	1	4	1	1	1	3	1	1	1	1	1	1	1	350	1	5	1	1	-1
11	2	3	3	2	3	4	3	3	3	4	3	2	2	450	0	7	-1	1	1
11	3	1	2	3	2	1	2	2	2	2	2	3	3	550	2	4	-1	-1	-1
11	4	2	4	4	4	2	4	4	4	3	4	4	4	650	3	6	1	-1	1
12	1	3	4	4	3	2	3	2	2	3	3	2	2	550	0	6	-1	1	1
12	2	1	3	3	2	4	4	3	4	2	2	4	4	650	3	4	1	1	-1
12	3	4	1	1	1	1	1	1	1	1	1	1	1	350	1	5	1	-1	1
12	4	2	2	2	4	3	2	4	3	4	4	3	3	450	2	7	-1	-1	-1
13	1	4	1	1	1	1	1	1	1	1	1	1	1	350	0	4	1	-1	1
13	2	3	2	3	2	4	2	2	4	4	3	2	4	450	3	6	-1	-1	-1
13	3	1	4	2	4	3	4	4	3	2	2	4	2	550	1	7	-1	1	1
13	4	2	3	4	3	2	3	3	2	3	4	3	3	650	2	5	1	1	-1
14	1	4	2	1	1	2	2	1	1	1	1	1	1	450	1	4	-1	1	1
14	2	3	3	3	3	3	3	3	2	3	2	2	2	550	0	6	1	1	-1
14	3	2	4	4	4	4	4	4	4	2	4	4	3	650	2	7	1	-1	1
14	4	1	1	2	2	1	1	2	3	4	3	3	4	350	3	5	-1	-1	-1
15	1	1	1	4	2	3	2	2	2	2	2	2	3	550	2	4	1	1	-1
15	2	2	4	3	3	4	3	4	3	4	4	4	2	650	1	6	-1	1	1
15	3	3	3	2	4	2	4	3	4	3	3	3	4	350	3	7	-1	-1	-1
15	4	4	2	1	1	1	1	1	1	1	1	1	1	450	0	5	1	-1	1
16	1	1	2	2	4	2	2	2	4	3	4	4	4	350	3	7	-1	1	1
16	2	4	1	1	1	1	1	1	1	1	2	1	3	450	2	5	1	1	-1
16	3	3	4	3	3	4	4	4	3	4	3	2	1	550	0	6	1	-1	1
16	4	2	3	4	2	3	3	3	2	2	1	3	2	650	1	4	-1	-1	-1

References

Ben-Akiva, M., Morikawa, T., Shiroishi, F., 1991. Analysis of the reliability of preference ranking data. Journal of Business Research 23(3), 253-268.

Brazell, J. D., Diener., C. G., Karniouchina, E., Moore, W. L., Séverin, V., Uldry, P. F., 2006. The no-choice option and dual response choice designs. Marketing Letters 17, 255-268.

Bunch, D. S., Louviere, J. J., Anderson, D., 1996. A comparison of Experimental Design Strategies for Multinomial Logit Models: The Case of Generic Attributes. Technical report. University of California, Davis. http://faculty.gsm.ucdavis.edu/bunch/.

Burgess, L., Street, D. J., 2003. Optimal designs for 2^k choice experiments. Communications in Statistics - Theory & Methods 32, 2185-2206.

Burgess, L., Street, D. J., 2005. Optimal designs for choice experiments with asymmetric attribute. Journal of Statistical Planning and Inference 134, 288-301.

Burgess, L., Street, D. J., Viney, R., Louviere, J. J., 2006. Design of choice experiments in health economics. In Jones, A. M. (Ed), The Elgar Companion to Health Economics, Chapter 39. Edward Elgar, Cheltenham.

Carson, R., Anderson, D., Arabie, P., Bunch, D., Hensher, D. A., Johnson, R., Kuhfeld, W., Louviere, J. J., Steinberg, D., Swait, J., Timmermans, H., Wiley, J., 1993. Experimental Analysis Of Choice. Marketing Letters 5, 351-368.

Chapman, R. G., 1984. An Approach To Estimating Logit Models of a Single Decision Maker's Choice Behavior. Advances in Consumer Research 11, 656-661.

Chapman, R. G., Staelin, R., 1982. Exploiting Rank Order Choice Set Data Within the Stochastic Utility Model. Journal of Marketing Research 19, 288-301.

de Palma, A., Myers, G. M., Papageorgiou, Y. Y., 1994. Rational Choice under an Imperfect Ability to Choose. American Economic Review 84, 419-440.

DeShazo, J. R., Fermo, G., 2002. Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency. Journal of Environmental Economics and Management 44, 123-143.

El-Helbawy, A. T., Bradley, R. A., 1978. Treatment Contrasts in Paired Comparisons: Large-sample Results, Applications and Some Optimal Designs. Journal of the American Statistical Association 73, 831–839.

Green, P. J., 1984. Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives. Journal of the Royal Statistical Society, Series B (Methodological) 46(2), 149-192.

Greene, W., 2003. Econometric Analysis, 5th ed. Prentice Hall, Upper Saddle River, NJ.

Hausman, J., Rudd, P, 1987. Specifying and Testing Econometric Models for Rank-Ordered Data. Journal of Econometrics 34(1-2), 83-104.

Heckman, J. J., 1979. Sample Selection Bias as a Specification Error. Econometrica 47(1), 53-161.

Heckman, J. J., 1990. Varieties of Selection Bias. The American Economic Review 80(2), 313-318

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

Horsky, D., Rao, M. R., 1984. Estimation of Attribute Weights From Preference Comparsons. Management Science 30(7), 801-822.

Huber, J., Zwerina, K., 1996. The Importance of Utility Balance in Efficient Choice Designs. Journal of Marketing Research 33, 307–317.

Huber, J., Train, K., 2001. On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths. Marketing Letters 12(3), 259-269.

Kanninen, B. J., 2002. Optimal designs for multinomial choice experiment. Journal of Marketing Research 39, 214-227.

Kessels, R., Goos, P., Vandebroek, M., 2006. A Comparison of Criteria to Design Efficient Choice Experiments. Journal of Marketing Research 43, 409-419.

Kuhfeld, W. F., Tobias, R. D., Garratt, M., 1994. Efficient Experimental Designs with Marketing Research Applications. Journal of Marketing Research 31, 545-557.

Louviere, J. J., 2001. What If Consumer Experiments Impact Variances As Well As Means: Response Variability As A Behavioral Phenomenon. Journal of Consumer Research 28, 506-511.

Louviere, J. J., 2004a. Complex Statistical Choice Models: Are the Assumptions True, and If Not, What Are the Consequences?. CenSoC Working Paper No. 04-002. Keynote Address, Discrete Choice Workshop in Health Economics, University of Oxford, 22 September.

Louviere, J. J., 2004b. Random Utility Theory-Based Stated Preference Elicitation Methods: Applications In Health Economics With Special Reference To Combining Sources of Preference Data. CenSoC Working Paper No. 04-001. Keynote Address, Australian Health Economics Society Conference, Canberra, ACT, 3 October.

Louviere, J. J., Eagle, T., 2006. Confound It! That Pesky Little Scale Constant Messes Up Our Convenient Assumptions. Proceedings, 2006 Sawtooth Software Conference, Sawtooth Software, Sequem, Washington, 211-228.

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

Louviere, J. J., Hensher, D. A., Swait, J., 1999. Conjoint analysis methods in the broader context of preference elicitation methods. In Gustafson, A., Hermann, A., Huber, F. (eds.), Conjoint Measurement: Methods and Applications, 279-318. Springer, Berlin.

Louviere, J. J., Hensher, D. A., Swait, J., 2000. Stated Choice Methods: Analysis and Application. Cambridge University Press, Cambridge.

Louviere, J. J., Meyer, R. J., Bunch, D. S., Carson, R. T., Dellaert, B., Hanemann, M., Hensher, D. A., Irwin, J., 1999. Combining Sources of Preference Data for Modelling Complex Decision Processes. Marketing Letters 10, 187-204.

Louviere, J. J., Street, D., Carson, R., Ainslie, A., DeShazo, J. R., Cameron, T., Hensher, D. A., Kohn, R., Marley, A. A. J., 2002. Dissecting the Random component of Utility. Marketing Letters 13, 177-193.

Luce, D., Suppes, P., 1965. Preference, Utility, and Subjective Probability. In Bush, R.R., Galanter, E. (eds.), Handbook of Mathematical Psychology, 249-410. John Wiley and Sons Inc., New York.

Magidson, J., Vermunt, J. K., 2007. Removing the Scale Factor Confound in Multinomial Logit Choice Models to Obtain Better Estimates of Preference. Proceedings, Sawtooth Software Conference, Santa Rosa, CA (October), forthcoming.

Pihlens, D. A., Louviere, J. J., 2004. How Indecisiveness in Choice Behaviour affects the Magnitude of Parameter Estimates obtained in Discrete Choice Models. Proceedings, ANZMAC Conference, Wellington, New Zealand.

Revelt, D., Train, K. E., 1998. Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. Review of Economics and Statistics 80, 1-11.

Richardson, A. J., 2002. Simulation Study of Estimation of Individual Specific Values of Time Using an Adaptive Stated Preference Survey. Paper presented at the Annual

Meetings of the Transportation Research Board, Washington, DC.
Sandor, Z., Wedel, M., 2001. Designing Conjoint Choice Experiments Using Managers' Prior Beliefs. Journal of Marketing Research 38, 430-444.
Sandor, Z., Wedel, M., 2002. Profile Construction in Experimental Choice Designs for Mixed Logit Models. Marketing Science 21, 455-475.
Séverin, V., Louviere, J. J., Finn, A., 2001. The Stability of Retail Shopping Choices Over Time and Across Countries. The Journal of Retailing 77, 185-202.
Shinghal, N., Fowkes, T., 2002. Freight mode choice and adaptive stated preferences. Transportation Research Part E: Logistics and Transportation Review 38(5), 367-378.
Street, D. J., Burgess, L., 2004a. Optimal and Near-Optimal Pairs for the Estimation of Effects in 2-level Choice Experiments. Journal of Statistical Planning and Inference 118, 185-99.
Street, D. J., Burgess, L., 2004b. Optimal stated preference choice experiments when all choice sets contain a specific option. Statistical Methodology 1, 37-45.
Street, D. J., Burgess, L., 2007. The Construction of Optimal Stated Choice Experiments: Theory and Methods. Wiley, Hoboken, New Jersey.
Street, D. J., Burgess, L., Louviere, J. J., 2005. Quick and Easy Choice Sets: Constructing Optimal and Nearly Optimal Stated Choice Experiments. International Journal of Research in Marketing 22, 459-470.
Street, A. P., Street, D. J., 1987. Combinatorics of Experimental Design. Oxford Science Publications, Clarendon Press, Oxford.
Swait, J. D., Adamowicz, W., 2001. Choice, Environment, Market Complexity and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice. Organizational Behavior and Human Decision Processes 86(2), 141-167.
Swait, J. D., Louviere, J. J., 1993. The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. Journal of Marketing Research 30, 305-314.
Thurstone, L. L., 1927. A law of comparative judgment. Psychological Review 34, 273-286.
Toubia, O., Simester, D. I., Hauser, J. R., Dahan, E., 2003. Fast Polyhedral Adaptive Conjoint Estimation. Marketing Science 22, 273-303.
Toubia, O., Hauser, J. R., Simester, D. I., 2004. Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis. Journal of Marketing Research 41, 116-131.
Train, K. E., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge.
Tversky, A., 1964. On the Optimal Number of Alternatives at a Choice Point. Journal of Mathematical Psychology 1, 386-391.
van Berkum, E. E. M., 1987. Optimal Paired Comparison Designs for Factorial Experiments. PhD thesis, Centrum voor Wiskunde en Informatica. CWI Tract 31.