

# Chapter 11

## Frontiers in Modeling Discrete Choice Experiments: A Benefit Transfer Perspective

Richard T. Carson, Jordan J. Louviere, John M. Rose  
and Joffre Swait

**Abstract** Given increasing survey costs, transferring model estimates obtained from one location and survey and applying them to another location is becoming increasingly appealing. The transfer of previously estimated model outputs to new application contexts has the potential to reduce the need for new large-scale data collection in the new application context as well as reduce the effort required to develop new models. As such, significant savings in cost and time can be achieved. Nevertheless, advantages in time and cost savings may be outweighed due to biases introduced if the transferred model does not adequately represent the behavior of individuals in the new application context. This chapter explores what benefits transfer means within the context of discrete choice experiments and outlines the challenges and possible improvements that could be made.

**Keywords** Benefits transfer · Discrete choice models · Discrete choice experiments · Challenges

### 11.1 Introduction

There is strong demand for benefit transfer (BT) in evaluating environmental policies because the number of benefit-cost assessments that need to be performed is large relative to the number of original benefit estimation studies. This is true

---

R.T. Carson  
University of California, San Diego, CA, USA  
e-mail: rcarson@ucsd.edu

J.J. Louviere (✉)  
University of South Australia Business School, Sydney, Australia  
e-mail: Jordan.Louviere@unisa.edu.au

J.M. Rose · J. Swait  
Institute for Choice, University of South Australia, Sydney, Australia  
e-mail: John.Rose@unisa.edu.au

J. Swait  
e-mail: Joffre.Swait@unisa.edu.au

despite the existence of many studies estimating various types of environmental benefits. The Environmental Values Resource Inventory (EVRI) database, maintained by Environment Canada in conjunction with several other countries including Australia, France, the United Kingdom and United States, contains over 3000 benefit estimates, and many more exist in the literature.<sup>1</sup> The demand for benefit transfer is obviously driven by the cost and time required to perform high-quality original benefit assessments. A BT exercise typically tries to infer the value of environmental policies whose outputs differ from those of a policy that was the subject of a formal assessment, either with respect to (a) the population of interest, (b) the time period of interest, and/or (c) one or more attributes of the policy valued in the original study(s).<sup>2</sup>

A major source of valuing information used in BT exercises comes from discrete choice experiments (DCEs).<sup>3</sup> This paper addresses a narrow question related to DCEs and BT; namely, the increasing use of more advanced statistical approaches to model DCE data, which have focused on incorporating various aspects of consumer heterogeneity. Estimates from studies using DCEs are used in several distinct ways in a BT context, so it is useful to start with a brief overview of that literature.

The seminal formulation of the issues involved in undertaking a BT exercise were put forth in a 1992 symposium in *Water Resources Research* (Brookshire and Neill 1992). The recent Johnston and Rosenberger (2010) review and chapters in this handbook examine many key issues involved in benefit transfer.<sup>4</sup> Over time a key distinction has evolved in the literature dealing with “value transfer” from a study [e.g., mean willingness to pay (WTP) for a particular policy], or “function transfer” from a study (Loomis 1992).<sup>5</sup> The function transfer approach has bifurcated into two distinct branches: one that uses a utility/valuation function estimated from a single or small number of related DCE studies (often in conjunction with other information such as demographic characteristics of the new location) to

---

<sup>1</sup><https://www.evri.ca/>.

<sup>2</sup>We use the word “policy” rather than “good” throughout to emphasize that most environmental goods are provided via some type of policy action. Differences in details of how a policy is implemented, including the perceived effectiveness of the government, can be important in carrying out a BT exercise.

<sup>3</sup>Louviere et al. (2000) provide the standard DCE treatment. Adamowicz et al. (1994) and Carson et al. (1990) provide initial DCE examples in the environmental literature. A recent review article focusing on environmental applications of DCE is Hoyos (2010). Carson and Louviere (2011) provide a common nomenclature for stated preference survey questions with an emphasis on environmental applications to try to clear up confusion over how various terms (including “DCE”) are used.

<sup>4</sup>Other overviews include Boyle et al. (2010), Navrud and Ready (2007) and Rosenberger and Loomis (2003). Rolfe and Bennett (2006) look at using DCEs in a benefit-transfer context.

<sup>5</sup>In both cases, the issue arises as to whether one study or multiple studies should be used for benefit transfer. Adjustments are often made to account for factors that differ between the original estimate and the new situation, and they sometimes also are made with the valuation function when changes to the variable values that appear in the valuation function are not thought to adequately capture differences between original and new situations.

provide an estimate for the new policy; and a second that combines a sizeable number of estimates from disparate individual studies, often based on very different methodologies via a formal statistical meta-analysis approach.

In our view, transferring values for a policy directly or using those values in a meta-analysis look similar; that is, any modeling approach that produces higher quality estimates for the original policy should result in higher quality BT transfer estimates. It is far less obvious that estimation of a more advanced choice model capable of more flexibly estimating consumer heterogeneity will produce higher quality outputs that are more reliable when being inferred (i.e. transferred) to a different context. In part, this is due to the well-recognized statistical issue that highly parameterized models often fit in-sample data quite well, but produce lower quality out-of-sample estimates than simpler models. An example where consumer heterogeneity was modeled in an initial study and then transferred to a new situation that indicated the more complex model was more successful in transferring BT to the new situation is provided by Colombo et al. (2007). The generalizability of this result and how much it depends on the specific aspects of consumer heterogeneity modeled remain open research questions.

DCEs are being used more frequently in empirical environmental valuation applications; so they provide many of the more current benefit estimates and functions used in BT exercises. Further, DCEs that explicitly value policy attributes can substantially expand the range of policies that can be valued relative to contingent valuation (CV) surveys that typically value only one scenario. The latter property greatly expands the range of what attribute-based DCEs can be used to do in a BT, but not necessarily their accuracy. For example, Kaul et al. (2011) show in their meta-analysis of BT exercises that CV studies focused on valuing single policies produced lower average transfer errors than those derived from more complex attribute based DCEs.<sup>6</sup> Thus, it may be better to think of a DCE with a substantial number of attributes and levels as more of a direct competitor to a meta-analysis where all of the estimates come from a single valuation technique focused on one type of policy. Like most meta-analyses, the emphasis is not on valuing one policy scenario but rather a substantial range of policy scenarios.

Morrison et al. (2002) provided a pioneering empirical example of using a DCE for BT. They implemented DCEs in three surveys involving two Australian wetlands (Gwydir and Macquarie), and interviewed in two different locations (Moree, a rural area, and Sydney). The Sydney population was interviewed separately about both wetlands, so that it was possible to look at benefit transfers to different

---

<sup>6</sup>The policies being valued in a BT exercise are not the same, so one should not draw strong conclusions about the relative performance of different techniques as the “source” of differences in BT exercises. More direct comparisons clearly are needed to make a more informed judgment. In general, however, what likely occurs is a tradeoff: the analyst can construct a model with more parameters involving policy attributes that allows for predictions (without ad hoc adjustments) to a broad range of policy changes, but at some cost to the accuracy of those predictions. This may be due to fitting a larger number of parameters, greater reliance on functional form assumptions, and/or less comprehensive depiction of individual attributes and their levels, given the same survey length.

locations and different populations. Respondents were asked to answer five choice sets that each had three alternatives [the first was a status quo (SQ) alternative]. Morrison et al. (2002) fit a conditional logit model with alternative specific constants for the two hypothetical alternatives and included various interactions of respondent demographics with the choice of a non-status quo alternative. They compared implicit prices estimated from the three surveys as well a set of nine randomly chosen policies for which compensating surplus was calculated from a BT perspective, using the mean levels of the demographic variables at the target site. Tests suggested that many of the implicit prices did not differ statistically across the surveys, but there were some clear exceptions. They found that the average mean difference in model estimates was 32 % across the nine policy transfer scenarios, consistent with good quality transfers in the literature. Transfers for different wetlands using the same population involved less error than transfers for the same wetland using different populations.

Jiang et al. (2005) and Rolfe and Bennett (2006) provided other early BT tests using DCEs. These and other earlier benefit transfer applications were based on conditional and nested logit models and illustrate that there are a number of different comparisons that can be of policy interest, such as the ranking of policy options in addition to the usual implicit prices and WTP for particular attribute bundles. Colombo et al. (2007) provided the first BT comparison for a random parameters mixed logit model (Train 2009), which was used to capture stochastic heterogeneity in consumer preferences. They conducted two parallel surveys focusing on two similar policies involving soil erosion policies in two different regions of southern Spain, using them to predict each other's results. Like the Morrison et al. (2002) study, the Colombo et al. (2007) BT exercise should represent an ideal context for BT. They compared BT errors for 27 policy scenarios, randomly chosen from the full factorial, using a conditional logit specification and mixed logit specifications with and without correlated parameters.<sup>7</sup>

Colombo et al. (2007) found that the mixed logit specification without correlated errors produced smaller average transfer errors across the 27 transfer scenarios than the mixed logit model with correlated errors, which, in turn, dominated transfers from the conditional logit model specification.<sup>8</sup> On average the transfer error was 38 % lower using the uncorrelated mixed logit model than the conditional logit model, but there were cases where the simple conditional logit model dominated. Surprisingly, we could not find other benefit transfer tests in the literature using mixed logit or other ways to model consumer heterogeneity. Colombo et al. (2007) were careful to note that more tests would be required before one could rely on

---

<sup>7</sup>The study had six attributes, four of which were assumed normally distributed and two that were fixed, including cost. Similar to Morrison et al. (2002), demographic variables were interacted with an ASC and demographics from the target site were used in the transfer.

<sup>8</sup>Transfer errors in percentage terms were substantially larger on average in Colombo et al. (2007) than in Morrison et al. (2002), illustrating that a myriad of factors are likely at work beyond the particular valuation technique or modelling strategy used.

empirical regularities to determine the relative influence of different approaches to modelling DCE data on the performance of subsequent BT exercises.<sup>9</sup>

Many researchers have wanted to move beyond the workhorse conditional logit model (McFadden 1974) for one of two reasons.<sup>10</sup> The first reason is the independence of irrelevant alternatives (IIA) property that underlies the conditional logit model, making it computationally tractable even with modest computer resources. However, this comes at the cost of imposing strong restrictions on the pattern of substitution relationships between alternatives. The second is an interest in modeling different types of consumer heterogeneity. Initially researchers pursued models that directly relaxed the IIA property like nested logit, which is still quite popular, and multinomial probit, which for various reasons has never seen substantial applied use. For the purpose of modeling DCE data, nested logit models were first used in modeling environmental choice data (e.g., Carson et al. 1990) as a way to deal with the fact that the SQ alternative often behaves quite differently than other alternatives.

Although the use of nested logit models is still fairly common, recent work has put much more emphasis on relaxing the assumption that all consumers are identical except for a draw from the same error distribution. The motivation for this work is twofold: (1) relaxation of IIA, and (2) development of models that more realistically capture differences in consumers. There are three distinct ways to incorporate consumer heterogeneity. The first is to assume that at least some consumers have different preference parameters. Initially, this was achieved by interacting one or more attribute parameters with characteristics of consumers, effectively allowing different types of consumers to have different parameters, which accounts for systematic sources of heterogeneity. Although this often provides useful insights into consumer preferences, it does not allow for continuous (or discrete) distributions of preference parameters, nor does it directly attack the IIA issue.

More recently the random parameters logit [popularly known as a mixed logit (Train 2009)] has become the most popular approach for modeling consumer preference heterogeneity. This approach allows for a range of consumer preference parameters that are assumed to follow some distribution, typically normal or log-normal. This model effectively assumes that individual consumers follow a conditional logit model; hence each consumer is assumed to adhere to the IIA property. The mixed logit model is one way of dealing with IIA violations at the aggregate level if a violation is driven by the assumption that all consumers have the same preference parameters. Additionally, it can also deal with IIA at the individual level via an error-in-variables approach that defines shared stochastic effects through categorization dummies (e.g., private vs. public modes of transport). These dummies reflect IIA violations arising from shared unobserved attributes at the alternative level (see McFadden and Train 2000).

---

<sup>9</sup>There have been subsequent BT exercises using mixed logit (e.g., Baskaran et al. 2010), but these do not seem to have systematically tested the performance of variants of mixed logit versus conditional logit models.

<sup>10</sup>Hensher et al. (2005) provide a general overview of the properties of different choice models.

The second approach is to recognize that consumers may be characterized by draws from different error distributions, with each distribution having different scale parameters, reflecting a form of heteroscedasticity. This heteroscedasticity may be due to several factors including differential ability to choose between alternatives or differences in the importance of unobserved variables on choice behavior. The latter leads to scale heterogeneity models. The third approach is to assume that there are pure types of consumers and that sample members can be represented by these pure types (or mixtures of them), leading to what are popularly known as latent class models. It is possible to combine these different types of consumer heterogeneity, such that one can have models with both preference and scale heterogeneity as well as latent class models with preference and/or scale heterogeneity. However, as we discuss later, such models may not be well-identified statistically and can experience computational difficulties. In the extreme, one can have “individual”-level models whereby one estimates a model specific to each respondent in a DCE (Frischknecht et al. 2014).

Initial moves away from conditional logit models, such as nested logit, treated observations on choices as independent in the sense of observing only one choice per individual or, if more than one choice was observed, not linking those choices together. Models currently at the choice modeling research frontier exploit and often require multiple choice observations from the same individual under different conditions. This is typically accomplished using DCEs with multiple choice sets, but in principle revealed preference (RP) data of this type can also be collected (e.g., Swait et al. 2004). An example of this would be to have individuals record their fishing trips in a diary format. If one matches this with other data sources on temporal differences in fishing quality (crowding and other factors), one would have repeated choice occasions from the same individual under different conditions. If one has multiple choice occasions per individual, one can examine whether choice behavior changes across choice sets. The standard framework used assumes no systematic change across choice sets, which greatly simplifies statistical identification of key parameters. However, a number of empirical tests suggest that this assumption does not generally hold and several competing hypotheses have been put forward that predict specific types of changes.

Much of the work using more advanced choice models is associated with DCEs because the ability to control choice stimuli can greatly facilitate estimating more complex models. Consequently, we also discuss the role of experimental design for estimating advanced choice models. Another recent advance is the ability to collect extra preference information in each DCE choice set beyond that of the most preferred alternative. Thus, we also discuss a variant of best-worst preference elicitation that elicits most and least preferred choices within a set of alternatives (see for example Louviere et al. 2008; Marley et al. 2008). In addition to obtaining repeated stated preference (SP) observations, one can combine SP and RP data to obtain multiple choice occasions for an individual. This naturally raises questions about differences in “scale” in the two choice contexts (Swait and Louviere 1993), and ways to capture the differences and/or take them into account.

In the remainder of the chapter, we provide an overview of different ways to model consumer heterogeneity with an eye toward using these models in BT exercises. The starting point for our discussion is to lay out the behavioral and statistical assumptions underlying the conditional logit model that serves as a reference model for much applied work. Most of the issues we address later can be shown to follow from applying the same model to all choices and all individuals. Then we turn to a discussion of several advanced models currently in use, focusing on how each relaxes one or more key assumptions of the conditional logit model, and how they relate to each other. A critical issue in thinking about these models from a BT perspective is the role that covariates play (usually demographics that can be observed both at the donor site where the original study was performed and the target or transfer site). Next, we discuss experimental designs used to collect choice data and the impact they can have on the types of models that can be estimated and the precision of the associated parameter estimates. After that discussion, we consider combining different types of data: e.g., combining data from multiple DCEs in appropriate ways that can substantially extend the BT beyond that of the original models.

## 11.2 Behavioral and Statistical Framework Underlying the Conditional Logit Model

The foundational framework for econometric analyses of choosing one among  $J$  objects or alternatives in a choice set  $S$  is the utility-maximizing consumer. For each discrete object  $j \in S$ , say, a recreation site, the consumer  $n$  forms a judgment/evaluation/utility measure:

$$U_{nsj} = U(X_{nsj}, Z_n | \beta_{nsj}, \theta_{nj}). \quad (11.1)$$

where  $X_{nsj}$  is a vector of quality attributes for the alternative including the price to “consume” it,  $Z_n$  is a vector of socio-demographic characteristics (e.g., income, age, gender) and other consumer specific information (e.g., attitudes) describing the individual,  $\beta_{nsj}$  is a deep vector of preference parameters associated with  $X_{nsj}$  and  $\theta_{nj}$  is a vector of parameters associated with  $Z_n$ . Parameter vectors  $\beta_{nsj}$  and  $\theta_{nj}$  may be constrained in various ways, such as being made generic across alternatives and fixed across respondents. Utility function (1) arises from a Lancasterian framework (Lancaster 1966), whereby consumers define preferences on the basis of benefits (termed “characteristics” by Lancaster) generated by the attributes of the good in question.<sup>11</sup> Our representation of the utility function above directly connects

---

<sup>11</sup>The Lancasterian model collapses downward to the standard framework used in most micro-economic theory if all goods are represented by only an alternative specific constant (ASC) and price.

attributes and prices to the overall judgment, and is therefore a simplified representation. In the final act of the decision process, the decision maker is assumed to select alternative  $i^* \in S$  such that  $U_{nsi^*} \geq U_{nsj}$  for all  $j \neq i^*$ . Note that this description is deterministic in nature.

The process description we made above is based on some critical assumptions, some of which we mention below.

1. The decision maker is aware of and uses all relevant information in making her judgment about the attractiveness of an alternative.
2. Further, relevant information is available for all alternatives.
3. The decision maker is exhaustive in her evaluations, not only in terms of information use, but also in terms of “looking at” all alternatives in  $S$ . This is done irrespective of the number of alternatives in  $S$  and of the cost of performing evaluations.
4. Selection of the preferred alternative is done by ranking all alternatives according to their utility, and choosing the highest valued alternative.

Although making these assumptions explicit may be unfamiliar and seem unnecessary, we do this to remind the reader that the underlying behavioral decision process adopted in extant models of choice, particularly all those to be examined in this paper, depicts decision makers as fully rational, all-knowing, inexhaustible utility maximizers. No matter how sophisticated the econometric approaches employed in the models we discuss, these (and other, very important) assumptions are [always] present.

Bringing the analyst into the mix requires us to introduce the possibility that the analyst does not have the full knowledge set available to the decision maker, and so we must allow for a stochastic component to utility. We rewrite Eq. 11.1 below to reflect this:

$$U_{nsj} = V(X_{nsj}, Z_n | \beta_{nsj}, \theta_{nj}) + \varepsilon_{nsj}(\mu_n) \quad (11.2)$$

where  $V(\bullet)$  is the deterministic component of utility (i.e., that part of total utility known to the decision maker that an analyst can specify), whereas  $\varepsilon_{nsj}(\mu_n)$  is the stochastic utility that accounts for the difference between the decision maker’s total utility and the deterministic component known to the analyst and  $\mu_n$  is a deep parameter of the stochastic utility. It is perhaps more enlightening to rewrite Eq. 11.2 as follows to emphasize what actually gives rise to stochastic utility in random utility theory:

$$\varepsilon_{nsj}(\mu_n) = U_{nsj} - V(X_{nsj}, Z_n | \beta_{nsj}, \theta_{nj}). \quad (11.3)$$

It is in this sense that we can say that “analyst ignorance” about the underlying total utility gives rise to stochastic utility, and hence, to probabilistic (as opposed to



deterministic) choice. The analyst’s key relationship between the observed choice of alternative  $i \in M$  and the total utility construct Eq. 11.2 is this expression:

$$P_{nsi} = P\{U_{nsi} \geq U_{nsj}\} \quad \text{for all } i \neq j, i, j \in S, \tag{11.4}$$

where  $P_i$  is the probability that  $i \in S$  is chosen, all other quantities as previously defined.

The crucial step in operationalizing expression Eq. 11.4 into different models of choice involves the stochastic specification of the vector  $\varepsilon = (\varepsilon_{nsi}(\mu_n), \dots, \varepsilon_{nsj}(\mu_n))$ . As is well known, the workhorse Multinomial Logit (MNL) model (Ben-Akiva and Lerman 1985)

$$P_i = \exp(\mu_n V(X_{nsi}, Z_n | \beta_{nsi}, \theta_{ni})) / \sum_{j \in S} \exp(\mu_n V(X_{nsj}, Z_n | \beta_{nsj}, \theta_{nj})) \tag{11.5}$$

results if we assume that the elements of  $\varepsilon$  are independent and identically distributed (IID) Type I Extreme Value random variables, that is,

$$\varepsilon_{nsj}(\mu_n) \sim F(w) = \exp(-\exp(-\mu_n w)), \quad -\infty < w < \infty, \mu \geq 0, \forall j \in S. \tag{11.6}$$

Ben-Akiva and Lerman (1985) and Hensher et al. (2005) also provide detailed discussions associated with the derivation of Eq. 11.5. It is important to consider two critical properties of model Eq. 11.5:

1. A confound exists between the deterministic ( $V$ ) and stochastic utility components (parameter  $\mu_n$ ) because of the inseparable and multiplicative nature between the two parts: these always show up in the form  $\mu_n V$ .
2. As a result of the IID assumption, the MNL model has the Independence of Irrelevant Alternatives (IIA) property. This implies that the odds of choosing one alternative over another are influenced only by their own utilities, but not influenced by the utilities of other alternatives. This property can lead to counterintuitive implications about behavior in empirical contexts. To make the IIA assumption and its implications explicit, consider taking the ratio of the probabilities for two competing alternatives (often referred to as the odds ratio):

$$\begin{aligned} \frac{P_{nsi}}{P_{nsh}} &= \frac{\left( \exp(\mu_n V(X_{nsi}, Z_n | \beta_{nsi}, \theta_{ni})) / \sum_{j \in S} \exp(\mu_n V(X_{nsj}, Z_n | \beta_{nsj}, \theta_{nj})) \right)}{\left( \exp(\mu_n V(X_{nsh}, Z_n | \beta_{nsh}, \theta_{nh})) / \sum_{j \in S} \exp(\mu_n V(X_{nsj}, Z_n | \beta_{nsj}, \theta_{nj})) \right)} \\ &= \frac{\exp(\mu_n V(X_{nsi}, Z_n | \beta_{nsi}, \theta_{ni}))}{\exp(\mu_n V(X_{nsh}, Z_n | \beta_{nsh}, \theta_{nh}))} \end{aligned} \tag{11.7}$$

As can be seen from Eq. 11.7, only the utility expressions of the two competing alternatives being considered actually matter; the utility functions of all other alternatives present in  $S$  drop out. Models such as the MNL force this property for all pairs of alternatives; however, issues arise when decision makers are more or less likely to substitute alternative  $i$  or  $h$  with another alternative  $j \in S$ , and hence the ratio of the probabilities are not independent of the presence or absence of other alternatives as implied by Eq. 11.7. In the following sections, we examine models that relax these characteristics, but in this section we discuss them from the perspective of BT.

To take these in order, let us first consider the consequences of the confound between the scalar  $\mu_n \geq 0$  and the deterministic utility  $V$ . As first pointed out by Ben-Akiva and Lerman (1985), for a given set of observations and context, they are inseparable. Between contexts  $c_1$  and  $c_2$  (say, original measurement and transfer application contexts) two sources of differences can occur: scalars  $\mu_{c_1}$  and  $\mu_{c_2}$  (we drop subscript  $n$  here as it is common practice to assume that differences between  $\mu$  are due to context effects rather than differences between consumers) differ, and/or  $V_{jc_1}$  and  $V_{jc_2}$  differ. That is, in a benefits transfer exercise one may not “transfer” successfully because (a) stochastic utility distributions differ between contexts, (b) systematic utility functions differ, or (c) both stochastic and systematic utilities differ. Swait and Louviere (1993) proposed a basic statistical test to determine if  $\mu_{c_1} = \mu_{c_2}$  conditional on the assumption that  $V_{jc_1} = V_{jc_2}$  for all  $j$ ; if systematic utility functions differ between contexts, benefits transfer is a moot question. If both components differ between contexts, the confound between stochastic and systematic utilities implies that there is no way to determine why preferences do not transfer. This discussion makes it clear that at a basic level the deck is stacked against BT.

With respect to the IIA property of the MNL model, this is often thought to be too strong a theoretical restriction to hold in general. On the one hand, it should be noted that whether or not IIA is a reasonable assumption to impose on observed choices is an empirical issue. Additionally, IIA is an individual-level property that may or may not hold in aggregate choice data, whether or not it is reasonable to hold for any particular individual. A certain dataset may contain choices that display IIA-like behavior. On the other hand, choice model forms exist (e.g., nested logit, multinomial probit; see Swait 2006) that allow for flexibility in representing interdependence in substitution patterns. To make our discussion with respect to benefits transfer more concrete, consider this nested logit formulation:

$$P_{nsi} = \left( \frac{\exp(\mu_1 V_{nsi})}{\sum_{j \in C_{(i)}} \exp(\mu_1 V_{nsj})} \right) \left( \frac{\exp(\mu C_{(i)})}{\sum_{h=1}^H \exp(\mu_h)} \right), \quad \forall i \in S, \quad (11.8a)$$

$$I_h = \frac{1}{\mu_1} \ln \left( \sum_{j \in C_h} \exp(\mu_1 V_{nsj}) \right) \quad (11.8b)$$

where the set  $S$  is subdivided into  $H$  nests/clusters  $C_h$  ( $C_{(i)}$  is the cluster to which alternative  $i \in S$  belongs), which are collectively exhaustive and have no elements in common,  $\mu_1 \geq 0$  is a scalar to be explained,  $I_h$  is a nest-specific inclusive value measure that is known to be the expectation of the maximum utility of the alternatives in  $C_h$ , and all other quantities are previously defined. This model arises if one assumes that the  $\varepsilon$ 's are correlated within clusters  $C_h$  and uncorrelated between clusters instead of assuming that the stochastic utilities are IID Type I Extreme Value distributed. Specifically, the joint cumulative distribution function of the  $\varepsilon$  is given by

$$F(\varepsilon) = \exp(-G(\exp(-\varepsilon_{ns1}), \dots, \exp(-\varepsilon_{nsJ}))), \quad -\infty < \varepsilon < \infty, \quad (11.9a)$$

$$G(w_{ns1}, \dots, w_{nsJ}) = \sum_{h=1}^H \left( \sum_{j \in C_h} w_{nsj}^{\mu_1} \right)^{\mu/\mu_1}, \quad w_{nsj} \geq 0, j = 1, \dots, J. \quad (11.9b)$$

This stochastic assumption allows us to specify the correlation between stochastic components as follows (Swait 2006):

$$\rho_{ih} = \begin{cases} 1 - \left(\frac{\mu}{\mu_1}\right)^2 & \text{if } C_{(h)} = C_{(i)} \\ 0 & \text{if } C_{(h)} \neq C_{(i)} \end{cases} \quad i, h \in S. \quad (11.10)$$

Conceptually, this stochastic definition of the behavior of the  $\varepsilon$ 's results in a covariance matrix for these random variables that is homoscedastic (equal diagonal terms, since  $\mu$  applies to all alternatives), with non-zero covariances for all alternative pairs sharing cluster membership and zero off-diagonal terms for all pairs not sharing clusters. A model like multivariate probit has a more general covariance matrix,<sup>12</sup> but it will conceptually arrive at the same point of making choice probabilities exhibit a type of non-IIA responses. It is obvious that the nested logit model can capture non-IIA behavior if we form the odds ratio for alternatives  $h$  and  $i$  using Eq. 11.8a, which we find to depend upon whether or not the pair of alternatives share a cluster:

$$\frac{P_i}{P_h} = \frac{\exp(\mu_1 V_i)}{\exp(\mu_1 V_h)} \cdot \left\{ \begin{array}{ll} 1 & \text{if } C_{(h)} = C_{(i)} \\ \frac{\exp(\mu C_{(i)})}{\exp(\mu C_{(h)})} & \text{if } C_{(h)} \neq C_{(i)} \end{array} \right\} \quad i, h \in S \quad (11.11)$$

Thus, this straightforward extension to the MNL model allows one to avoid the IIA property empirically, but imposes preference homogeneity within clusters.

Now, let us consider the implication of this more general model with respect to BT. Clearly, if the structure of an IIA violation, i.e. the nests (or stochastic utility correlation structure) in measurement and transfer application contexts differ

---

<sup>12</sup>Bunch (1991) provides identification restrictions.

significantly between contexts, it may not be possible to successfully transfer between them. This effect is over and above that arising from the basic confound between stochastic and systematic utilities within each context. Even if the systematic utilities are identical across contexts, if the clustering structure is too dissimilar (e.g., it is place- or time-specific) or the quantity  $(\mu/\mu_1)$  is small (i.e. correlation is high, imposing pairwise requirements on stochastic utility for transfer to occur), BT may induce large errors, which may well be a key underlying difficulty with BT exercises that not seem to be clearly appreciated.<sup>13</sup>

### 11.3 Advanced Choice Models Available for Use with DCE Data

In this section, we move from a consideration of the nature of the underlying utility function to econometric issues that must be resolved for BT by examining how the estimates to be transferred arise. Of course, these are linked, but here we emphasize issues related to modeling data collected from samples of the population of interest where various sources of heterogeneity are considered. We focus on the general case, the likelihood function of discrete choice models, rather than examine all possible models that can be estimated.

Typically, the parameters  $\beta$  associated with each utility function  $V_{nsi}$  are unknown and must be estimated from data. Let  $y_{nsi}$  equal one if  $i$  is the chosen alternative in choice situation  $s$  shown to respondent  $n$ , and zero otherwise. In other words,  $y$  represents the outcomes of a discrete choice experiment. The parameters can be estimated by maximizing the likelihood function  $L$ ,

$$L = \prod_{n=1}^N \prod_{s \in S_n} \prod_{j \in J_{ns}} (P_{nsi})^{y_{nsi}}, \quad (11.12)$$

where  $N$  denotes the total number of respondents and  $S_n$  is the set of choice situations faced by respondent  $n$  and  $P_{nsi}$  is a choice probability. This choice probability is expressed as equation Eq. 11.5 for the conditional logit or MNL model and equation Eq. 11.8a for the nested logit model.

The log-likelihood functions of more advanced models can also be derived with a simple substitution of the appropriate choice probability for the model being estimated in Eq. 11.12. For example, in the mixed logit model, the parameters to be estimated are structural parameters representing the population moments of some underlying (multivariate) distribution (e.g., the mean(s) and standard deviation(s) of

---

<sup>13</sup>Note that transferring values via a meta-analysis does not solve this problem. The same good can be worth different amounts in different contexts where the difference between the contexts is not observed. The best that a meta-analysis can do in this instance is to average over a small number of values for the same good derived in different contexts.

a (multivariate) normal distribution). To estimate the parameters of the mixed logit model using simulated maximum likelihood involves taking draws from the random parameter distribution(s), calculating the expected or average logit choice probability over the simulation draws, and substituting this into equation Eq. 11.12. However, the underlying likelihood function is the same.

Rather than maximize the likelihood function, it is more common to maximize the log of the likelihood function because the product of a series of probabilities typically produces extremely small values that most computing software cannot adequately handle. Taking the logs of the probabilities produces large negative values, which when multiplied, produce even larger negative values. Consequently, the log-likelihood function of the model, shown below, is typically preferred:

$$LL = \ln \left[ \prod_{n=1}^N \prod_{s \in S_n} \prod_{j \in J_{ns}} (P_{nsi})^{y_{nsi}} \right]. \tag{11.13}$$

If one assumes that the choice observations are independent over both respondents and choice situations and one uses the mathematical properties  $\ln(n_1 n_2) = \ln(n_1) + \ln(n_2)$  and  $\ln(n_1)^{y_{nsi}} = y_{nsi} \ln(n_1)$ , and applies the same mathematical rules to choice tasks,  $s$ , and alternatives  $J$ , one can rewrite equation Eq. 11.13 in the more commonly recognized form:

$$LL = \sum_{n=1}^N \sum_{s \in S_n} \sum_{j \in J_{ns}} y_{nsi} \ln(P_{nsi}). \tag{11.14}$$

It is possible for some advanced models, such as the mixed logit model, to relax the assumption that responses are independent within a respondent. In this case, the log-likelihood function of the model becomes:

$$LL = \sum_{n=1}^N \ln \left( E \left( \prod_{s \in S_n} \prod_{j \in J_{ns}} (P_{nsi})^{y_{nsi}} \right) \right), \tag{11.15}$$

where  $E \left( \prod_{s \in S_n} \prod_{j \in J_{ns}} (P_{nsi})^{y_{nsi}} \right) = E(P^*) = \int_{\beta} P_n^*(\beta) f(\beta|\theta) d\beta$ ,  $P^*$  is the probability of the chosen alternative, and  $f(\beta|\theta)$  is the multivariate probability density function of  $\beta$ , given the structural parameters  $\theta$ .

This chapter deliberately examines discrete choice models from the perspective of the log-likelihood function for two primary reasons. First, it is important to understand that the basics of the estimation process are fundamentally the same for all discrete choice models, and are independent of any particular assumed model specification. The reason for this is that if one views the log-likelihood function as described above, the logic can be extended to any discrete choice model by substituting the specific probability for the model of interest. In turn, this allows for

a more general discussion of the relevant issues without getting bogged down in discussions of the specifics of all model forms. Second, a focus on the log-likelihood makes it clear that, regardless of model type, the estimation process involves nothing more than trying to find the parameter estimates (whether random or fixed) that maximize the probabilities over choice tasks for the alternatives observed to have been chosen in the data. The reason for the latter statement is that only the chosen alternative matters in the log likelihood function (i.e., if  $y_{nsi} = 0$ , the function is zero) and mathematically as  $P_{nsi} \rightarrow 1$ , the log of  $P_{nsi} \rightarrow 0$ , suggesting that as the log likelihood function approaches zero, the estimated parameters best predict the chosen alternatives.<sup>14</sup>

The latter point is particularly important, and so deserves more attention. As suggested by Eq. 11.4, the choice probabilities associated with any choice model are a function of the utilities specified by the analyst, which in turn are potentially a function of the design attributes,  $X_{nsi}$ , including the price to “consume” the alternative, socio-demographic characteristics of respondents, potential attitudes respondents have about the object(s) of interest, the decision context associated with the choice made, and the parameters to be estimated,  $\beta$ . The vast majority of DCE studies typically consider the design attributes, but generally do *not* consider socio-demographic characteristics. Even fewer consider attitudes and/or decision context(s).

Even if one tries to consider all possible decision influences, there are many possible ways that one can enter them into a final model specification. For example, one could enter socio-demographic variables linearly into the alternative specific constants (ASCs) of one or more utility functions (e.g., with the SQ alternative) and/or one can interact them with one or more of the design attributes. For more advanced models like latent class models, socio-demographic characteristics may enter the model as part of a class assignment model (Swait 1994), whereas in the case of nested logit one can extend Eq. (11.8b) to specify the inclusive value measure as a function of exogenous variables outside of any contained in  $V_{nsi}$ . Similar specification possibilities arise if one has attitudinal and/or decision context measures.

Thus, how one specifies the utility function of the model matters greatly from an estimation perspective, whether one is interested in BT or not. Unfortunately, the degree to which this matters is an empirical issue, which may vary from context to context. At a simple level, if a salient variable is omitted from the analysis, covariances between parameter estimates can adversely affect the parameter outcomes for those variables that entered the utility function. Unfortunately, one can learn about this only by specifically and systematically testing this effect by including all available variables into the model in all possible ways. Further, if interactions exist and are not incorporated into the model specification, then one runs a risk of introducing endogeneity bias into the estimation process.

---

<sup>14</sup>It should be noted though that the probabilities depend on all of the alternatives.

To elaborate, discrete choice models of all types assume that  $V_{nsi}$  and  $\varepsilon_{nsi}(\mu)$  are orthogonal to one another. If one omits a salient interaction term involving one or more variables in the model specification, the observed component and the error term of the model become correlated. For example, if one estimates the parameters of a model where price *should* enter as both a main effect and an interaction in the model, but price *only* enters as a main effect, the price variable is associated with both the observed component and error term. In such cases, the parameters of the affected variables captured in the modeled component of utility will likely be biased, and in the immediately preceding price example, the price parameter will be biased. Naturally in such cases, all implicit prices also will be biased, and the amount of endogeneity bias can differ for different parameters, depending on where the problem resides in the utility function. Such problems are not limited to DCEs, and they are not limited by data collected in surveys. For example, if different respondents make different assumptions about a decision context and/or treat ambiguous attributes differently, or respondent attitudes are related to the true decision process but not observed, endogeneity bias may arise and require sophisticated methods to detect and deal with it.

The preceding discussion raises a very important, fundamental question for BT. That is, *before* one considers transferring outcomes from one location (or data source) to another, one must know if what is being transferred is correct (i.e., unbiased), regardless of the specific model type estimated. More specifically, one cannot ask what model type will provide the best outcomes for BT unless the estimates for *all* model types compared are based on the best representations (most correct approximations) of the underlying decision processes. To do otherwise is to merely compare the robustness of various model forms to violations similar to those discussed above. In all likelihood, the latter exercise merely poses an empirical question, the answer to which will depend on the type and degree of violation experienced.

Additionally, from a BT perspective the preceding discussion illustrates other broader issues that one faces when trying to transfer estimates obtained from one location (or data source) to another. For example, if data are available on all relevant decision variables in two different sites (data sources)—whether the data represent DCE design attributes, socio-demographic characteristics, attitudes or information about decision contexts—the first question one should ask is whether the same set of decision variables should enter the utility specifications of models that would be independently estimated from the two data sets. If the answer to the first question is “yes,” one should then ask if they enter the utility functions of both models in *exactly the same way*. The latter differs slightly from the issue raised in the preceding paragraph because one is asking not only if the utility specification for the estimation site (data source) is correct, but also if it best represents the specification at the transfer site (data source). If the two utility specifications differ, then one needs to consider the particular effects or calculations being transferred, and whether such differences are likely to matter.

For example, if implicit prices are being transferred, problems can arise if there are interactions between design attributes and other exogenous variables and/or if

different data transformations should be applied to variables entering the model. To illustrate this problem, consider the implicit price for the following utility specification ( $x_c$  the price,  $x_k$  a quality attribute):

$$V = \cdots + \beta_1 x_k + \beta_2 x_k x_c + \beta_3 x_c^2 + \cdots \quad (11.16)$$

The marginal implicit price for a one unit increase in  $x_k$  becomes

$$WTP_k = -\frac{\Delta x_c}{\Delta x_k} = -\frac{\frac{d}{dx_k}(\beta_1 x_k + \beta_2 x_k x_c + \beta_3 x_c^2)}{\frac{d}{dx_c}(\beta_1 x_k + \beta_2 x_k x_c + \beta_3 x_c^2)} = -\frac{\beta_1 + \beta_2 x_c}{\beta_2 x_k + 2\beta_3 x_c}. \quad (11.17)$$

Note that the value of the exogenous variable  $x_k$  does not drop out of the calculation; so, for the case of BT, one needs to know the value of the exogenous variable at the transfer site (data source) in addition to the parameters being transferred. An additional issue can arise in cases where one requires utility specifications that are highly non-linear in the attributes. In this case, one or more attributes in the estimation model may be irrelevant to the transfer site (data source); hence, one must carefully attend to what precisely is being transferred. Moreover, leaving aside the issue of models that allow for consumer heterogeneity in the parameter estimates, one also may need to consider the issue of heterogeneity (or range) in the  $X$ s associated with BT. Given the seriousness of these issues and the uncertainty likely to be associated with any particular BT application, one may be tempted to rely on simple models, such as simple linear in the parameter and linear in the attribute utility specifications. Unfortunately, however, our earlier discussion on potential biases that can occur when one misspecifies the true decision process highlights the fact that there clearly are potential risks in doing this.

Putting all the above aside, let us now consider a model specification in which both preferences, represented as vector  $\beta$ , and scale, represented by a scalar  $u_n$ , are assumed to vary over the sampled population, such that  $V = u_n \beta$ . It is possible to rewrite  $V = \alpha_n = u_n \beta$ , where the elements in  $\alpha_n$  must be correlated as each term in  $\beta$  is multiplied by a common scalar  $u_n$ . When viewed in this light, the confound between scale and preference suggests that what is being modeled is also a form of correlation. If one or more parameters in  $\alpha_n$  are fixed, or if all parameters are assumed to be randomly distributed but uncorrelated, then an analyst is making the [implicit assumption] that scale is homogenous across the sample. If one treats all parameters as random and correlated (via, say, a Cholesky decomposition of the parameter covariance matrix) in the mixed logit model, that model also allows for random scale. The question then becomes what is being transferred in BT exercises when one uses these advanced models.

Ideally, the model from the estimation site should capture both scale and preference heterogeneity, which implies allowing for correlated random parameters. Yet, one now must transfer not only mean estimates but also the entire covariance structure of random parameter terms, and assume that all these terms are similar to those at the transfer site, an important assumption that may not hold empirically.



To further complicate the BT issue, up to this point we have been deliberately vague about the specifics of how to appropriately model attitudinal data for discrete choice models. Typically, attitudinal data is collected using multi-attribute likert scales or some other similar type of approach. One often observes researchers entering such consumer-reported value(s) as an independent variable in the utility specification, but such an approach can be problematic. As is well known, consumers use rating scales differently, such that a rating of (say) two means different things to different consumers (see for example Lee et al. 2007). Thus, this is a type of measurement error in which the true underlying attitudes are measured by some latent unobserved value. More recently, Rungie et al. (2011) provide comprehensive statistical theory to integrate structural equation models and choice models, providing a theoretically appropriate way to incorporate latent variables in the latter. Despite these advances in incorporating latent constructs, such as attitudes in models, new issues arise for BT, as analysts now must transfer parameter estimates for the design attributes and parameters for the latent attitudinal variables. The latent variables now must be imputed for the transfer site somehow. In turn, this suggests that our earlier discussion on endogeneity also applies.

We end this particular discussion with an interesting and likely case-specific question; namely, what if the best models for two separate sites (data sources) are *not* the same. That is, what would be the likely outcome of a BT exercise if the model that best represented the decision processes at the estimation site has one particular model form (e.g., a mixed logit model), while the model that best fits the true decision processes at the transfer site (data source) is another model form (e.g., conditional or nested logit)? This would seem to be an important future research question that begs several other questions that, in turn, suggest that researchers need to consider other possible models, decision processes and even the prospect that different people use different decision processes.<sup>15</sup>

## 11.4 Experimental Design and Collection of SP Data for Estimating Choice Models

Thus far our discussion of models for choice data has been quite limited in the sense that data satisfying the conditions necessary for estimating such models have been assumed to be available. Data from DCEs typically come from stated preference surveys due to greater flexibility in controlling the stimuli seen by individuals, which in turn minimize certain statistical issues in estimating the desired model and

---

<sup>15</sup>It is important to note that none of these issues are avoided by using estimated WTP from DCEs for particular goods as inputs to the meta-analyses rather than making the transfer based directly on the utility function estimated using DCE data. That is because all the specification issues discussed as well as issues involving scale, can have substantial influence on WTP estimates from both the original study and estimates derived for BT exercises.

provide rigorous tests of particular hypotheses.<sup>16</sup> Typically, SP data result from some formal statistical design process, but there are exceptions. For example, some researchers still use random draws from the total possible sets of choice sets to construct SP surveys. To the extent that the number of draws is sufficiently large this will approximate the population of choice sets, but as is the case with any random sampling procedure, there can be serious departures from representativeness in small samples and the statistical identification of particular parameters may be tenuous (Carson et al. 2009).

SP data collection processes for DCEs can be specifically designed to maximize the power of tests of particular hypotheses and/or minimize estimation errors associated with specific parameters of particular models. Of course, there is no free lunch because SP studies face issues of external validity, due in no small part to incentive compatibility issues, hypothetical situations and a wide array of design-related and/or induced effects on outcomes.

### *11.4.1 Single Binary Discrete Choice CV Experiments*

Contingent valuation (CV) methods have been used for over five decades, with a recent book citing thousands of studies (Carson 2012). This chapter views the standard single binary discrete choice (SBC) question recommended by Arrow et al. (1993) and used in many high profile CV studies (e.g., Carson et al. 2003) as the simplest special case of a DCE, whereby the choice options are limited to an SQ alternative and a substantive policy alternative. One attribute, typically the payment cost, is randomly varied. Payment costs are chosen to help identify the shape of the underlying distribution of WTP and minimize the confidence interval around key statistics like mean or median WTP. Generally speaking, there is no need for the values of the payment vehicle to be chosen randomly from a range. The number of payment amounts (often referred to as bid values) and their values depend critically on three factors: (a) the distributional assumption made about WTP, (b) prior information assumed about the value of the parameters of that distribution, and (c) the statistic(s) of interest. In the extreme case where a two-parameter distribution for WTP has been assumed and the values of the distribution's parameters are known with certainty, the typical result is that only two bid values should be used. Assumptions of more flexible WTP distributions or allowing for considerable uncertainty over parameter values generally results in more bid values being optimal, but optimal designs typically result in the use of four to eight bid values.

The SBC format allows one to trace out the WTP distribution in the population of interest if a parametric assumption about the WTP distribution is made (Cameron

---

<sup>16</sup>RP data also can be collected with DCEs in a laboratory setting, field tests or other contexts, such as internet websites selling products where it is possible to control and randomly vary stimuli seen by individuals. In this section, we assume that a DCE is used to collect SP data as that is by far the most common application.

1988; Cameron and James 1987) or one uses a nonparametric step-function to approximate it (Carson and Steinberg 1990; Haab and McConnell 2002). The SBC format with appropriate auxiliary conditions has desirable incentive properties (Carson and Groves 2007), but these properties arise mainly because the SBC format collects so little information from each respondent. For all practical purposes, preference parameters of individual respondents are unidentified beyond the interval in which their WTP lies, although some parameters related to observable individual characteristics (e.g., being an environmentalist) can be estimated at the average of members of the group. Technically, it is possible to estimate the difference in the population's WTP for two policies that differ by two levels of a single attribute by asking two statistically equivalent subsamples SBC questions that differ only on that dimension. This is known as an external scope test (Arrow et al. 1993), but it quickly gets prohibitively expensive when extended to multiple samples. The desire to value change in multiple attributes that individually have multiple levels pushes one in the direction of asking respondents to make choices in more than one choice set. Moving to DCEs with multiple choice sets also greatly facilitates more detailed modelling of consumer heterogeneity.

### *11.4.2 DCE with Multiple Choice Sets*

DCEs were pioneered by Louviere and Hensher (1983) and Louviere and Woodworth (1983). DCEs evolved by drawing on work from several sources, including conjoint measurement (Luce and Tukey 1964), information integration theory (e.g., Anderson 1970), discrete multivariate analysis of contingency (crosstab) tables (e.g., Bishop et al. 1975), probabilistic discrete choice models (e.g., McFadden 1974) and the design of statistical experiments for linear models (e.g., Box et al. 2005). Basically a DCE is a sparse, incomplete contingency (crosstab) table. The table is constructed: i.e., it is purposefully designed a priori so that certain statistical properties are achieved. Typically, these properties include identification of certain effects associated with a particular statistical specification of research interest, the precision with which these effects should be estimated, and/or other properties. Statistical design theory provides theory and construction methods for achieving these purposes.

There are two basic types of DCE designs, namely generic and alternative-specific. Generic designs (GDs) are used when the choice options are unlabeled (e.g., Option A, Option B, etc.), and alternative-specific designs (ASDs) are used when the choice options are labeled (e.g., Manly Beach or Bondi Beach). Most of the advances in optimum design theory work for DCEs in the past decade has focused on GDs. Little work is available for ASDs outside of the transportation literature (although see Rose and Bliemer 2009). As noted below, the vast majority of DCE applications have used GDs; however, in many (but not all) cases, the researchers should have used ASDs.

A GD can be seen as a way to construct choice sets offering options described by a generic set of attributes and associated levels; such designs can be used to construct pairs, triples, etc., of options. Typically, environmental applications of DCEs present a status quo option and two or more competing options. So, the design problem is to select the “best” way to configure the choice sets to satisfy certain statistical criteria of interest. Typically, the criteria are identification (ensuring that all parameters of an assumed statistical model of interest can be estimated) and precision (the estimated parameters are minimum variance estimates). There can be other criteria, such as trying to maximize the fit of the model in- and/or out-of-sample. The onus is on the designer to insure that ALL relevant attributes are included (to avoid omitted variable bias), and that the levels are appropriate (ranges not too large or too small, qualitative levels correct and relevant). Louviere and Woodworth (1983) and Louviere et al. (2000) discuss these designs for cases where a constant option is present in each choice set (e.g., status quo).

Alternative-specific designs are not frequently seen in environmental applications, and may pose issues for BT. That is, to the extent that the modeling results from ASDs are specific to particular labeled options, the model parameter estimates and/or calculations based on them may not transfer to sites not included in the estimation set. The references cited in the preceding paragraph also discuss ASDs. Basically, such designs can be constructed by using designs for linear models as long as there is a constant option in each choice set, such as not choosing any of the options. The advantages of the design construction methods discussed in the cited references are that they allow one to estimate more general models and test violations of IIA.

Since 2000 there has been a rapidly growing literature in the design of DCEs. This literature has gone in three different directions. The first was an effort to examine the statistical properties of designs used in DCEs. This included efforts to improve the efficiency of different types of logit models, including the conditional logit (e.g., Bunch et al. 1996), the nested logit (Bliemer et al. 2009) and the mixed logit model (Bliemer and Rose 2010; Sándor and Wedel 2002), under various assumptions about the unknown parameters of the utility function, and also to study the implications of more complex models from a design perspective. The second was how to collect more information about preference in each choice set faced by respondents. The third was to ask whether particular designs used in DCEs influenced the choice behavior observed (see for example Bliemer and Rose 2011; Louviere et al. 2008).

Although orthogonal designs have been the traditional mainstay for DCEs and will continue to be so for many years to come, a number of researchers have begun to query the appropriateness of orthogonal designs for use in DCEs. Generally, the argument against the use of orthogonal designs is that the property of orthogonality may run counter to many of the desirable properties of the econometric models typically used to analyze DCE data (i.e., logit and probit models). Instead of merely looking at the correlation between the attribute levels, as with orthogonal designs, the latest theories related to the construction of experimental designs for DCEs seek to find designs that are statistically as efficient as possible in terms of predicted standard errors of the parameter estimates. Essentially, these designs try to maximize the information available from each choice situation and minimize the

standard errors of the estimated parameters. Unlike linear models, the asymptotic variance-covariance (AVC) matrix of discrete choice models is derived by taking the negative inverse of the expected second derivatives of the log-likelihood function of the model (see for example McFadden 1974). Given that the log-likelihood function is itself a function of the choice probabilities which are in turn a function of the parameter estimates, it is necessary for the analyst to provide priors to determine the statistical efficiency of a design before the design used in the field.

Different researchers have made use of different prior parameters over the years, with the two main types being the use of locally optimal priors (fixed parameters; e.g., Huber and Zwerina 1996) or Bayesian priors (distributions of potential prior parameter estimates; e.g., Sándor and Wedel 2001). Within the literature dealing with locally optimal priors, there also exist several different research streams, including those that assume non-zero parameter values (e.g., Carlsson and Martinsson 2002) and those that assume that the parameters values are all zero (e.g., Street and Burgess 2007). It is interesting to note that for this latter design class, the choice probabilities will be  $1/J$ . In this case, the logit model will approximate a linear model and an orthogonal design will be optimal. The inverse case is also true; an orthogonal design is akin to assuming that the parameters will be zero.

Marley et al. (2008) suggested that previous work in Best-worst Scaling (e.g., Marley and Louviere 2005) could be extended to DCEs as a way to collect extra choice information about the choice options in each choice set. Specifically, they proposed using the order information from best and worst (most preferred and least preferred) choices in each choice set to expand the available data to additional sets of implied choices. Expanding (or exploding) the choices is based on Luce and Suppes (1965), and was applied to Random Utility Theory (RUT)-based choice models by Beggs et al. (1981) and Chapman and Staelin (1982) and others. Louviere et al. (2008) show how to use the extra order information to estimate models for single individuals. More recently, Louviere (2014) discusses the underlying ideas in much more detail, explicitly linking them to earlier work in DCEs, and Frischknecht et al. (2014) discuss new ways to estimate models for single individuals that ensure accurate parameter recovery and convergence. Thus, we now have theory and methods for modeling single individuals using DCEs.

About five years ago, researchers in the Centre for the Study of Choice (CenSoC) in Australia noticed what seemed to be an unusually high number of experimental participants who exhibited deterministic choices, typically always choosing one level of one particular attribute (e.g., “always choose lowest price”) in Street and Burgess (2007) designs. Further investigation led to the discovery that this was a widespread phenomenon. This led them to design 64 different DCEs to rigorously test differences in implied preferences between types of designs for DCEs, numbers of attributes, numbers of choice sets, and numbers of attribute levels. Results are emerging as we write this, but we can say that there is evidence for relatively high proportions of deterministic choices (30–40 %) in Street and Burgess (2007) and Statistical Analysis System (SAS) designs (Kuhfeld 2010), but virtually no evidence of this in the Balanced Incomplete Block Design (BIBD) approach proposed by Louviere et al. (2008) or in randomly constructed sets of

choice sets. There also are systematic differences in attribute effects for five and eight attributes, with these effects frequently associated with quantitative attributes such as payment vehicles. Unfortunately, there is a growing trend for researchers not to test the results of DCE models against RP data, which is why this phenomenon may have gone undetected for some time. A clear and compelling research agenda emerging from these results is that not only models but types of designs need to be tested against RP data so that we can begin to understand which design strategies will produce the most accurate and reliable results for which purposes and contexts.

Closely related to the above comments is the fact that (as noted earlier) there seems to be far too much reliance on generic designs (Louviere et al. 2000, Chap. 4) in empirical research with DCEs. Indeed, in many cases, the appropriate design should be an alternative-specific design, but researchers seem to copy one another, and many researchers, particularly in applied economics (but also in marketing) routinely use generic designs. Generic designs are appropriate when one wants to generalize one's results to a generic class of options. However, researchers often focus on a specific class of problems, such as visits to particular competing recreational sites and/or particular competing treatments for a special health condition. Thus, it is fair to say that many researchers should be designing and implementing alternative-specific DCEs instead of generic DCEs.

Moreover, it is much harder to validate generic DCEs against RP data because RP data typically are associated with choices of particular competing options, such as products on store shelves in marketing applications, or transport modes in transport. Alternative-specific problems necessarily are narrower than generic problems, with the focus being on a particular set of choice options of interest. This set can include "any other options" and/or an outside good, but one rarely sees these included in DCEs. Alternative-specific designs have the decided advantage that they can simulate the features of real markets to any desired degree of accuracy required. That is, one can use such designs (literally) to create a variety of choice contexts such as store shelves or competing recreational opportunities.

Choice models for alternative-specific problems, not surprisingly, may or may not require alternative-specific utility functions, whereby  $J-1$  of the competing options has its own utility function with potentially different parameters. The  $J$ th option must be set equal to some constant for identification purposes. Louviere and Woodworth (1983) proposed that the latter option be the choice of none of the competing options, or another constant option like a status quo good. Such models can be specified as latent class models or random coefficient models, but one rarely sees this in practice (however, see Swait 1994).

From the standpoint of BT problems, one can clearly see the strong attraction of generic DCEs. Moreover, applied economics generally has tended to favor models without ASCs where possible, which contrasts strongly with marketing applications in which ASCs often are brand labels. Thus, a key question from a BT perspective is whether a donor site is something that can be characterized as a bundle of generic

attributes or whether it is unique in ways that influence BT.<sup>17</sup> Clearly, this is an issue that also influences other approaches to BT, like meta-analysis. The DCE approach illustrates the specific technical nature of assumptions made in using generic designs and how it can bias transferred valuation estimates.

## 11.5 Issues in Combining Data from Multiple Sources

It is possible to combine data from multiple sources to improve the performance of individual models and extend the ability of a particular dataset to make predictions to additional situations. There has been a long standing interest in combining SP and RP data (Adamowicz et al. 1994; Ben-Akiva and Morikawa 1990; Cameron 1992). Moving from RP data to SP data was seen as a way to overcome the limitations that RP data have of dealing only with a narrow range of attribute levels and potentially having very highly correlated attributes, but it was immediately obvious that there was useful information in both types of data, so it was natural to try to combine them. Originally, RP data were seen as the “gold” standard representing the ideal measure of the behavior one wanted to predict, but this perspective has become more nuanced over time.

RP data are often collected by government agencies, non-profit organizations and commercial firms. Examples include trips taken for recreation purposes and purchases of different types of goods. A key issue to note is that most RP data is collected in surveys. RP data are subject to the general reporting error issues involved in collecting information in surveys, plus some special considerations like memory recall effects. Another source of potential divergence in behavior predicted by RP and SP data that one must be aware of arises from the nature of samples taken relative to the population of interest, and one frequently observes quite different selection bias between surveys collecting the two types of data.<sup>18</sup>

What is by now well-known is that there may be considerable differences in scale between RP and SP and that there are ways for taking this into account (Swait and Louviere 1993). The RP and SP choice environments may be characterized by very noisy levels, even though all or most of the parameter estimates are proportional between the two types of data (Louviere et al. 2000, Chap. 13; Swait et al. 1994).<sup>19</sup>

Calibration of SP estimates to predict RP choices seems clearly merited if two conditions hold. The first is that the information available in the RP environment

---

<sup>17</sup>Burgess et al. (2012) provide an initial investigation of this question, asking whether different types of landscape configurations can be adequately represented as a bundle of attributes.

<sup>18</sup>One way to avoid this issue is to collect both types of data in the same survey, but this may exacerbate recall issues, particularly if respondents are asked about behavior in the more distant past.

<sup>19</sup>Divergence from proportionality, when it occurs, tends to be concentrated in the ASC and cost parameters.

and the noise that characterizes it represent the context in which one wishes to predict choice behavior. This may not always be the case, particularly for many environmental decisions or new products in marketing, where one wants to predict the behavior of informed consumers.<sup>20</sup> The second condition occurs if the incentive structure of the SP question is inconsistent with truthful preference revelation, which is most likely to be the case for private and quasi-public goods.<sup>21</sup> Although RP data for pure public goods in the desirable incentive context of a coercive payment mechanism do not generally exist, RP data sometimes exist with voluntary payment mechanisms. This presents a difficult situation in which the incentive structure of the RP context should lead to well-known free-riding behavior. For SP data, Carson and Groves (2007) show that it is optimal to over-pledge in the survey to encourage the undertaking of the fundraising effort and then to free-ride on the actual effort. This discussion suggests that while there can be considerable benefit to combining RP and SP data, such data-pooling should be done with careful consideration of the processes generating both types of data.

It is also possible to combine SP data sets. While there is little experience with doing this, the principles are largely the same as combining RP and SP datasets. There need to be two or more common pairs of choice alternatives across the two datasets. Differences in scale between the two datasets will have to be controlled for in a manner similar to RP and SP data. The ability to combine multiple SP data sets opens the possibility of systematically collecting and merging SP datasets with an eye toward being able to value an increasing array of environmental goods in a consistent manner.<sup>22</sup>

## 11.6 Conclusions

This chapter represents an initial foray into many questions related to how different modeling strategies influence BT error rates. We noted that particular attention should be paid to the role of how observable differences in donor and target sites are handled in the transfer exercise. This includes determination of relevant population subgroups that occur across the two sites. In turn, this suggests considerable potential could be realized by reweighting the data at the donor site to match the

---

<sup>20</sup>Participants in DCEs that collect SP data often are given considerably more information than they are likely to have had in an RP context, where purchase decisions are infrequent or made only once.

<sup>21</sup>Carson and Groves (2007) provide a neo-classical framework for examining the incentive structure of various SP elicitation formats in different contexts. For new products, incentives in a SP context tend to encourage over-estimation of the propensity to purchase, whereas for existing products incentives may exist to understate the propensity to purchase.

<sup>22</sup>For instance, a government agency could adopt a plan to undertake a new SP study valuing various environmental goods every year where a small number of choice alternatives were common across years.



target site on key characteristics and then re-estimating the original model with reweighted data.<sup>23</sup>

A general difficulty with most empirical BT tests is that they are designed to eliminate many messy details involved in transfers actually performed for policy purposes. From a DCE perspective, a key difficulty is that attributes and levels may not match well between donor and target sites. It may also be that some demographic or attitudinal information collected in the survey is unavailable at the target or that a considerable length of time has passed between the original study and the BT exercise. No doubt some or all of these factors can impact the quality of the BT. A key research question is whether quality suffers differentially with different modeling strategies, which in turn points to the need to develop a more comprehensive theoretical and econometric framework for transferring the information contained in DCEs to help inform policy decisions. In the latter case the use of DCEs for direct BT is much less developed than the meta-analysis approach, but clearly has the potential to provide a coherent framework for BT. There is much conceptual work to be done, and empirical comparisons across donor and target sites using a range of modelling approaches and outside data are clearly necessary.

## References

- Adamowicz, W., Louviere, J. J., & Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management*, 26, 271–292.
- Adamowicz, W., & Swait, J. (2012). Are food choices really habitual? Integrating habits, variety seeking and compensatory choice in a utility maximizing framework. *American Journal of Agricultural Economics*, 95, 17–41.
- Anderson, N. H. (1970). Functional measurement and psychophysical judgment. *Psychological Review*, 77, 153–170.
- Arrow, K., Solow, R., Portney, P., Learner, E., Radner, R., & Schuman, H. (1993). Report of the NOAA panel on contingent valuation. *Federal Register*, 58, 4601–4614.
- Baskaran, R., Cullen, R., & Colombo, S. (2010). Testing different types of benefit transfer in valuation of ecosystem services: New Zealand winegrowing case studies. *Ecological Economics*, 69, 1010–1022.
- Beggs, S., Cardell, S., & Hausman, J. A. (1981). Assessing the potential demand for electric cars. *Journal of Econometrics*, 17, 1–19.
- Ben-Akiva, M., & Lerman, S. (1985). *Discrete choice analysis: Theory and application to predict travel demand*. Cambridge, MA: MIT Press.

---

<sup>23</sup>Demographic variables observable at both sites are the source of information for reweighting information. Other interesting alternatives are possible; for example, if it were known that information availability and decision makers (Hensher and Rose 2009; Swait et al. 2012) differed across the two sites and/or one knew the mix of decision rules (Adamowicz and Swait 2012) used by individuals, it would be possible to reweight on this source of heterogeneity across people. More generally, it might be possible to estimate individual level models at the donor site and then use some type of propensity score matching procedure to infer individual level utility functions at the transfer site.

- Ben-Akiva, M., & Morikawa, T. (1990). Estimation of switching models from revealed preferences and stated intentions. *Transportation Research Part A*, 24, 485–495.
- Bishop, S. E., Feinberg, S., & Holland, P. (1975). *Discrete multivariate analysis*. Cambridge, MA: MIT Press.
- Bliemer, M. C. J., & Rose, J. M. (2010). Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B*, 46, 720–734.
- Bliemer, M. C. J., & Rose, J. M. (2011). Experimental design influences on stated choice outputs: An empirical study in air travel choice. *Transportation Research Part A*, 45, 63–79.
- Bliemer, M. C. J., Rose, J. M., & Hensher, D. A. (2009). Constructing efficient stated choice experiments allowing for differences in error variances across subsets of alternatives. *Transportation Research Part B*, 43, 19–35.
- Box, G. G., Hunter, W. G., & Hunter, J. S. (2005). *Statistics for experimenters: Design, innovation, and discovery* (2nd ed.). New York: Wiley.
- Boyle, K. J., Kuminoff, N. V., Parmeter, C., & Pope, J. C. (2010). The benefit-transfer challenges. *Annual Review of Resource Economics*, 2, 161–182.
- Brookshire, D. S., & Neill, H. (1992). Benefit transfers: Conceptual and empirical issues. *Water Resources Research*, 28, 651–655.
- Bunch, D. S. (1991). Estimability in the multinomial probit model. *Transportation Research*, 25B(1), 1–12.
- Bunch, D. S., Louviere, J. J. & Anderson, D. A. (1996). A comparison of experimental design strategies for Multinomial Logit Models: The case of generic attributes. Working paper, Graduate School of Management, University of California at Davis.
- Burgess, D., Finney, G. & Matthews, D. (2012, September). Landscape valuation: Choice experiments or contingent valuation. Paper presented at the 14th Annual International BIOECON Conference on Resource Economics, Biodiversity Conservation and Development, Cambridge, UK.
- Cameron, T. A. (1988). A new paradigm for valuing non-market goods using referendum data: Maximum likelihood estimation by censored logistic regression. *Journal of Environmental Economics and Management*, 15, 355–379.
- Cameron, T. A. (1992). Combining contingent valuation and travel cost data for the valuation of nonmarket goods. *Land Economics*, 68, 302–317.
- Cameron, T. A., & James, M. D. (1987). Efficient estimation methods for closed-ended contingent valuation surveys. *Review of Economics and Statistics*, 69, 269–276.
- Carlsson, F., & Martinsson, P. (2002). Design techniques for stated preference methods in health economics. *Health Economics*, 12, 281–294.
- Carson, R. T. (2012). *Contingent valuation: A comprehensive bibliography and history*. Northampton, MA: Edward Elgar.
- Carson, R. T., & Groves, T. (2007). Incentive and information properties of preference questions. *Environmental and Resource Economics*, 37, 181–210.
- Carson, R. T., Hanemann, W. M., & Steinberg, D. (1990). A discrete choice contingent valuation estimate of the value of Kenai King salmon. *Journal of Behavioral Economics*, 19, 53–68.
- Carson, R. T., & Louviere, J. J. (2011). The state of the art of environmental valuation with discrete choice experiments. *Environmental and Resource Economics*, 49, 539–559.
- Carson, R. T., Louviere, J. J., & Wasi, N. (2009). A cautionary note on designing discrete choice experiments: A comment on Lusk and Norwood’s “effect of experimental design on choice-based conjoint valuation estimates”. *American Journal of Agricultural Economics*, 91, 1056–1063.
- Carson, R. T. & Steinberg, D. (1990). Experimental design for discrete choice voter preference surveys. In *1989 Proceedings of the Survey Methodology Section of the American Statistical Association*. Washington, DC: ASA.
- Carson, R. T., Mitchell, R. C., Hanemann, W. M., Kopp, R. J., Presser, S. & Ruud, P. A. (2003). Contingent valuation and lost passive use: Damages from the Exxon Valdez oil spill. *Environmental and Resource Economics*, 25, 257–286.

- Chapman, R. G., & Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, 19, 288–301.
- Colombo, S., Calatrava-Requena, J., & Hanley, N. (2007). Testing choice experiment for benefit transfer with preference heterogeneity. *American Journal of Agricultural Economics*, 89, 135–151.
- Frischknecht, B., Eckert, C., Louviere, J. J. & Geweke, J. (2014). Incorporating prior information to avoid complete separation problems in estimating discrete choice models. In S. Hess & A. Daly (Eds.), *Handbook of choice modelling*. London: Elsevier.
- Haab, T. C., & McConnell, K. E. (2002). *Valuing environmental and natural resources: The econometrics of nonmarket valuation*. Northampton, MA: Edward Elgar.
- Hensher, D. A., & Rose, J. M. (2009). Simplifying choice through attribute preservation or non-attendance: Implications for willingness to pay. *Transportation Research Part E*, 45, 583–590.
- Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). *Applied choice analysis: A primer*. New York: Cambridge University Press.
- Hoyos, D. (2010). The state of the art of environmental valuation with discrete choice experiments. *Ecological Economics*, 69, 1595–1603.
- Huber, J., & Zwerina, K. (1996). The importance of utility balance and efficient choice designs. *Journal of Marketing Research*, 33, 307–317.
- Jiang, Y., Swallow, S. K., & McGonagle, M. P. (2005). Context-sensitive benefit transfer using stated choice models: Specification and contingent validity for policy analysis. *Environmental and Resource Economics*, 31, 477–499.
- Johnston, R. J., & Rosenberger, R. S. (2010). Methods, trends and controversies in contemporary benefit transfer. *Journal of Economic Surveys*, 24, 479–510.
- Kaul, S., Boyle, K. J., Pope, K. J., Parmeter, C. F. & Kuminoff, N. V. (2011, June). *Robust meta-analysis using median-quantile and nonparametric regression procedures: Investigating the validity of benefit transfers*. Paper presented at the Annual Summer Meeting of the Association of Environmental and Resource Economists, Seattle.
- Kuhfeld, W. F. (2010). *Marketing research methods in SAS experimental design, choice, conjoint, and graphical techniques*. Cary, NC: SAS Institute.
- Lancaster, K. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74, 132–157.
- Lee, J. A., Soutar, G., & Louviere, J. J. (2007). Measuring values using best-worst scaling: The LOV example. *Psychology and Marketing*, 24, 1043–1058.
- Loomis, J. B. (1992). The evolution of a more rigorous approach to benefit transfer: Benefit function transfer. *Water Resources Research*, 38, 701–705.
- Louviere, J. J. (2014). Modeling single individuals: The journey from Psych lab to the app store. In S. Hess & A. Daly (Eds.), *The state-of the art and the state-of-the-practice*. London: Elsevier.
- Louviere, J. J., & Hensher, D. A. (1983). Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event. *Journal of Consumer Research*, 10, 348–361.
- Louviere, J. J., Hensher, D. A., & Swait, J. (2000). *Stated choice analysis: Methods and analysis*. New York: Cambridge University Press.
- Louviere, J. J., Islam, T., Wasi, N., Street, D., & Burgess, L. (2008). Designing discrete choice experiments: Do optimal designs come at a price? *Journal of Consumer Research*, 10, 348–361.
- 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.
- Luce, R. D., & Suppes, P. (1965). Preference, utility and subjective probability. In R. D. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 249–410). New York: Wiley.
- Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1, 1–27.
- Marley, A. A. J., Flynn, T. N., & Louviere, J. J. (2008). Probabilistic models of set-dependent and attribute-level best–worst choice. *Journal of Mathematical Psychology*, 52, 281–296.

- Marley, A. A. J., & Louviere, J. J. (2005). Some probabilistic models of best, worst, and best-worst choices. *Journal of Mathematical Psychology*, *49*, 464–480.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *In frontiers in econometrics* (pp. 105–142). New York: Academic Press.
- McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, *15*, 447–490.
- Morrison, M., Bennett, J., Blamey, R., & Louviere, J. J. (2002). Choice modelling and tests of benefit transfer. *American Journal of Agricultural Economics*, *84*, 161–170.
- Navrud, S., & Ready, R. C. (2007). *Environmental value transfer: Issues and methods*. Dordrecht: Springer.
- Rolfe, J., & Bennett, J. W. (Eds.). (2006). *Choice modelling and the transfer of environmental values*. Cheltenham, UK: Edward Elgar.
- Rose, J. M., & Bliemer, M. C. J. (2009). Constructing efficient stated choice experimental designs. *Transport Reviews*, *29*, 587–617.
- Rosenberger, R. S., & Loomis, J. B. (2003). Benefit-transfer. In P. A. Champ, K. J. Boyle, & T. C. Brown (Eds.), *A primer on nonmarket valuation* (pp. 445–482). Northampton, MA: Edward Elgar.
- Rungie, C. M., Coote, L. V., & Louviere, J. J. (2011). Structural choice modelling: Theory and applications to combining choice experiments. *Journal of Choice Modelling*, *4*, 1–29.
- Sándor, Z., & Wedel, M. (2001). Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research*, *38*, 430–444.
- Sándor, Z., & Wedel, M. (2002). Profile construction in experimental choice designs for mixed logit models. *Marketing Science*, *21*, 455–475.
- Street, D., & Burgess, L. (2007). *The construction of optimal stated choice experiments: Theory and methods*. New York: Wiley.
- Swait, J. (1994). A structural equation model of latent segmentation and product choice for cross-sectional, revealed preference choice data. *Journal of Retailing and Consumer Services*, *1*, 77–89.
- Swait, J. (2006). Advanced choice models. In B. Kanninen (Ed.), *Valuing environmental amenities using stated choice studies: A common sense approach to theory and practice* (pp. 229–294). Dordrecht: Springer.
- Swait, J., Adamowicz, V., & van Buren, M. (2004). Choice and temporal welfare impacts: Incorporating history into discrete choice models. *Journal of Environmental Economics and Management*, *47*, 94–116.
- Swait, J., & Louviere, J. J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, *30*, 305–314.
- Swait, J., Louviere, J. J., & Williams, M. (1994). A sequential approach to exploiting the combined strengths of SP and RP data: Application to freight shipper choice. *Transportation*, *21*, 135–152.
- Swait, J., Popa, M. & Wang, L. (2012, July). A cost-benefit model of consumer information management and choice. Working Paper, Centre for the Study of Choice, University of Technology, Sydney.
- Train, K. E. (2009). *Discrete choice models with simulation* (2nd ed.). New York: Cambridge Economic Press.