7 Stated choice experimental design theory: the who, the what and the why

John M. Rose and Michiel C.J. Bliemer

1 INTRODUCTION

Unlike most survey data where information on both the dependent and independent variables is captured directly from respondents, stated preference data where respondents make decisions based in hypothetical markets, of which stated choice (SC) data is a subset, is unique in that typically only the dependent variable is provided by the respondent. With the exception of covariate information which is often ignored in most analysis, the primary variables of interest, consisting of attributes and their associated levels, are designed in advance and presented to the respondent in the form of competing alternatives in SC studies. However, increasing evidence of both an empirical (for example, Bliemer and Rose, 2011; Louviere et al., 2008) and theoretical nature (for example, Burgess and Street, 2005; Sándor and Wedel, 2001, 2002, 2005) suggests that the allocation of the attribute levels over the experiment may impact upon the model outputs obtained, particularly when small samples are involved. As such, rather than simply randomly assigning the attribute levels shown to respondents over the course of an experiment, experimental design theory has traditionally been applied to allocate the attribute levels to the alternatives in some systematic manner.

A review of the literature, however, suggests that little consensus exists as to what specific experimental design theory, or aspects thereof, are appropriate for SC studies. Indeed, over the past three decades, research in the area of experimental design construction specifically related to SC studies has resulted in a number of what appear to be competing paradigms, many of which claim superiority in their approach. Unfortunately, claims of superiority require objective criteria, and over the years, different researchers have selectively chosen criteria that support their chosen paradigm to the detriment of competing methods.

The objective of this chapter is therefore twofold. First, it is argued that despite the disparate nature of the existent literature leading to multiple claims of methodological superiority, there exists only a single underlying theory related to the construction of SC experiments. The only thing that differs is the underlying assumptions made by various analysts. Secondly, in presenting this theory it is discussed how the various researchers in this field have actually been reliant on this theory, either implicitly or explicitly, but under very different sets of assumptions. It is these assumptions that define the different approaches in constructing SC experimental designs.

The remainder of this chapter is set out as follows. Section 2 discusses what exactly an experimental design is and why it is important. Section 3 outlines a number of decisions that are required prior to generating the experimental design. Section 4 then provides a discussion of the theory of experimental design as it relates to SC studies. Section 5 provides a selective historical overview of key research groups working in the area of experimental design theory for SC studies. Section 6 comprises brief concluding comments.

2 WHAT IS AN EXPERIMENTAL DESIGN?

Conceptually, an experimental design is simply a matrix of values that is used to determine what goes where in an SC survey. These values may be either numbers or labels depending on how the analyst wishes to relate the information of the experiment to the respondents. The values that populate the matrix represent the attribute levels that will be used in the SC survey, whereas the columns and rows of the matrix represent the choice tasks (alternatively called choice sets or choice situations depending on the literature cited), attributes (sometimes referred to as factors) and alternatives (referred to as profiles in the marketing literature and treatment combinations in the mainstream experimental design literature) of the experiment. The layout of the design matrix is typically presented in one of two ways. Some researchers set up the experimental design matrix such that each row represents a different choice task and each column a different attribute within the experiment (see for example, Bliemer and Rose 2009; Rose and Bliemer 2009). In such cases, groups of columns form different alternatives within each choice task. Other researchers, however, set up the design matrix such that each row of the matrix represents an individual alternative and each column a different attribute (see for example, Carlsson and Martinsson, 2003; Huber and Zwerina, 1996; Kanninen, 2002; Sándor and Wedel, 2001, 2002). In these cases, multiple rows are grouped together to form individual choice tasks. Independent of how the matrix is set out, the experimental design performs the same function; that being the allocation of attribute levels to choice tasks.

The distribution of the levels of the design attributes over the experiment, via the underlying experimental design, may play a large part in whether or not an independent assessment of each attribute's contribution to the observed choices of sampled respondents can be determined. Further, the allocation of the attribute levels within the experimental design will also impact upon the statistical power of the experiment, in so far as its ability to detect statistical relationships that may exist within the data. This ability is related to the sample size of the study, thus given a large enough sample the statistical power of an experimental design may not matter. Nevertheless, for the sample sizes more commonly used in practice, the ability to retrieve statistically significant parameter estimates may be compromised given the selection of a relatively poor design (sample size issues are discussed in section 5) or the need to detect particular effects such as preference heterogeneity using more complex or advanced econometric models that may require larger sample sizes than simpler model types. The experimental design chosen by the analyst may therefore play a significant role in SC studies.

3 CONSIDERATIONS BEFORE GENERATING THE DESIGN

For any SC study, there exist many experimental designs that can be constructed and used in practice. The aim for the analyst is to choose a particular method and construct the design. How best to do this depends upon many different considerations. In this section, we discuss a number of decisions that have to be made prior to the design being generated. As such, these decisions are independent of the design generation process used; however, they may influence this process significantly.

The first decision required of the analyst is to decide whether the experiment will be labelled (that is, the experiment uses alternatives), the names of which have substantive meaning to the respondent other than indicating their relative order of appearance, for example, the alternatives might be labelled Sherwood Forest, Cork Bark Forest, The Tongass, or *unlabelled* (for example, the names of the alternatives only convey their relative order of appearance, for example, Policy A, Policy B, Policy C). Aside from affecting what results can appropriately been derived for the study (for example, elasticities have no substantive meaning in unlabelled experiments), from the perspective of design generation, this decision is important as it might directly impact upon the type and number of parameters that can or will be estimated as part of the study. Typically, unlabelled experiments will involve the estimation of generic parameters only whereas labelled experiments may involve the estimation of alternative specific and/or generic parameter estimates, hence potentially resulting in more parameter estimates than with an identical, though unlabelled, experiment.

Alongside considerations as to how the alternatives are to be presented to respondents (that is, labelled or unlabelled), advance knowledge of the response mechanism to be used and the number of alternatives (including no choice or status quo alternatives) to be shown to each respondent is also crucial. To date, most discrete choice studies use a response mechanism that captures first preference rankings only via a 'pick one' response. Recent developments, however, have seen an increase in obtaining either partial or full rankings data via say a best/worst response (for example, Flynn et al., 2008). Aside from psychological issues surrounding various response mechanisms, the type of response and the number of alternatives shown may impact upon the dimensions of the design in terms of the minimum number of choice tasks that are required for estimation purposes.

Unlike most other data types where an observation typically represents information captured about a specific respondent or agent, in discrete choice data each alternative j represents a unique observation. This is because each alternative is observed to be chosen or not, hence providing information down to this level of detail. In grouping the alternatives together in choice tasks, there therefore exist $J - 1$ independent choice probabilities within each choice task S which will be estimated. As such, for first preference (pick one) tasks, the total number of independent choice probabilities obtained from any given design will be equal to $S \cdot (J - 1)$ with the maximum number of parameters, K, that can be estimated from that design having to be less than or equal to this number. However, when rankings data is captured, it becomes possible to rank explode the data such that new choice tasks are formed, consisting of subsets of the original J alternatives that are formed after iteratively removing previously more highly preferred alternatives (in terms of the rank r assigned). Instead of $J - 1$, the total number of modeled choice observations will now be equal to $\sum_{r=1}^{R}(J - r)$, where $r = 1, 2, \ldots, R$, is an index of the ranking up to the total number of rankings provided for a given choice task. For example, assume a respondent provides two rankings points (best and worst), hence $R = 2$. The maximum number of parameters that can then be estimated is then given as

$S \cdot \sum_{r=1}^{R}(J - r)$. If only the first preference is captured, $R = 1$ and if the full ranking is captured, $R = J - 1$. As can be seen in this discussion, the size of the design (the number of choice tasks generated) is therefore a function of the number of parameters to be estimated, the number of alternatives shown and how the response mechanism relates to the estimation process to be used.

The number of attribute levels used is also an important pre-design decision, as is how they are to be modelled. This is because this will influence the number of parameters that can be estimated. If non-linear effects are expected for a certain attribute, then more than two levels need to be used for this attribute in order to be able to estimate these non-linearities. Typically, dummy and/or effects coding is used in order to estimate these non-linear effects. In this case, the number of parameters needed for each attribute is the number of levels minus one. Hence, the more levels used, the higher the minimum number of choice tasks required will be, as each additional level will require an additional parameter to be estimated. However, if the attribute is a continuous variable, one can also decide to estimate linear effects with only a single parameter per attribute, even though in the design multiple levels are used. In this case, the minimum number of required choice tasks does not increase as the number of levels increases.

Another decision that is required prior to generating the design is whether attribute level balance will be imposed when generating the design. Attribute level balance occurs when each attribute level appears an equal number of times for each attribute over the design. We discuss this concept in greater detail later as various researchers have applied the concept of attribute level balance differently. However, such a choice is important as it too may also impact on the size of the final design. Also, mixing the number of attribute levels for different attributes may yield a higher number of choice tasks or alternatives required for a given design (due to attribute level balance). For example, if there are three attributes with 2, 3 and 5 levels, respectively, then the minimum number of choice tasks for an attribute level balanced design will be 30 (since this is divisible by 2, 3 and 5 without remainder). On the other hand, if one would use 2, 4 and 6 levels, then a minimum of only 12 choice tasks would be required. Therefore, it is often suggested not to mix designs with too many different numbers of attribute levels, or at least have all even or all odd numbers of attribute levels, if attribute level balance is a design criterion.

The attribute level range of quantitative attributes is another important decision that needs to be made before the experimental design can be determined. Research suggests that using a wide range (for example, \$1–\$6) is statistically preferable to using a narrow range (for example, \$3–\$4) as this will theoretically lead to better parameter estimates (that is, parameter estimates with a smaller standard error), although using too wide a range may also be problematic (see Bliemer and Rose, 2010a). The reason for this is that the attribute level range will impact upon the likely choice probabilities obtained from the design, which will impact upon the expected standard errors from that design (see section 4). Having too wide a range will likely result in choice tasks with dominated alternatives (at least for some attributes), whereas a too narrow range will result in alternatives which are largely indistinguishable. We have to emphasize that this is a pure statistical property and that one should take into account the practical limitations of the attribute levels, in so much as the attribute levels shown to the respondents have to make sense. Therefore, there is a trade-off between the statistical preference for a wide range and practical considerations that may limit the range.

The number of choice tasks, as previously mentioned, is bounded from below by the number of degrees of freedom, which is influenced by: the number of parameters to be estimated; the number of alternatives that will be shown to respondents; the response mechanism as well as by other considerations such as the number of choice tasks required to achieve attribute level balance. Additionally, the design type may restrict the number of choice tasks. A (fractional factorial) orthogonal design sometimes needs (many) more choice tasks than the minimum number determined by the number of degrees of freedom and attribute level balance, merely because an orthogonal design may not exist or may be unknown for these dimensions. A full factorial design has a predetermined number of choice tasks, only influenced by the total number of attributes and the number of attribute levels.

A final consideration when generating experimental designs of SC studies is what model type is likely to be estimated on data collected using the design. At a minimum, the model type may influence the number of parameters to be estimated which will influence the number of choice tasks required of the design. More complex discrete choice models may also require larger sample sizes to detect particular effects hidden with the data. While most research has focused on generating designs for the multinomial logit model, more advanced models such as the mixed logit model may require different design characteristics and larger sample sizes in order to detect heterogeneity that may exist within the data. Given this background, the next section discusses experimental design theory in detail.

4 EXPERIMENTAL DESIGN THEORY

Independent of the specific application area, experimental design theory has at its core two common objectives: (i) the ability to detect independently the effects of multiple variables on some observable outcome and (ii) improving the statistical efficiency of the experiment. In many cases, these two objectives are not in themselves independent, with designs that allow for an independent assessment of various variables on some dependent variable being the same as those that are considered to be statistically efficient.

Statistical efficiency in experimental design terms relates to an increased precision of the parameter estimates for a fixed sample size. Statistical efficiency therefore is linked to the standard errors likely to be obtained from the experiment (and to a lesser extent to the covariances), with designs that can be expected to (i) yield lower standard errors for a given sample size, or (ii) the same standard errors given a smaller sample size, being deemed more statistically efficient. Fortunately, as will be argued in section 5, while different criteria have been applied at various times to measure statistical efficiency, the underlying definition of statistical efficiency has remained constant, even if it has not been appropriately applied in all cases. Once it is understood that, independent of the specific problem being examined, experimental design theory (as applied in the case of the SC context) is concerned with the standard errors (and covariances) of the parameters obtained from models to be estimated from data collected using a generated design, it is clear that what is of prime consideration is the relationship between the design and the resulting model variance-covariance matrix (from whence the standard errors are derived).

As highlighted, experimental design theory originated in fields other than SC experiments and hence developed specifically to address models tailored to particular data types corresponding to the problems that were being addressed at the time. Indeed, the original theories dealt specifically with experimental problems where the dependent variable was continuous in nature. As such, the resulting design theory was developed specifically for models capable of handling such data hence much of the work on experimental design theory has concentrated on use of analysis of variance (ANOVA) and linear regression type models (see Peirce, 1876). From a historical perspective this has had a significant impact upon the SC literature. The original SC studies, unsurprisingly, were used to test specific economic effects, such as the existence of indifference curves (May, 1954; MacCrimmon and Toda, 1969; Mosteller and Nogee, 1951; Rousseas and Hart, 1951; Thurstone, 1931), and even more recently concentrated on introducing and promoting the benefits of the new modelling method and did not concentrate specifically on the issue of experimental design (see, for example, Louviere and Hensher 1983; Louviere and Woodworth, 1983). As such, these earlier works understandably borrowed from the early theories on experimental design without considering whether they were appropriate or not for use with models applied to such data. Over time, as is often the case in academia, the designs used in these earlier SC studies became the norm and have largely remained so ever since.

Sporadic research over the years, however, has looked at the specific problem of experimental designs as related to econometric models estimated on discrete choice data. In order to calculate the statistical efficiency of a SC design, Fowkes and Wardman (1988), Bunch et al. (1996), Huber and Zwerina (1996), Sándor and Wedel (2001) and Kanninen (2002), among others, have shown that the common use of logit models to analyse discrete choice data requires a priori information about the parameter estimates, as well as the final econometric model form that will be used in estimation, although the final design may be more or less robust to deviations of model specification (see Bliemer and Rose, 2010a). Specifically, information on the expected parameter estimates, in the form of priors is required in order to calculate the expected utilities for each of the alternatives present within the design. Once known, these expected utilities can in turn be used to calculate the likely choice probabilities. Hence, given knowledge of the attribute levels (the design), expected parameter estimate values and the resultant choice probabilities, it becomes a straightforward exercise to calculate the asymptotic variance-covariance (AVC) matrix for the design, from which the expected standard errors can be obtained. The AVC matrix of the design, Ω_N, can be determined as the inverse of the Fisher information matrix, I_N, which is computed as the negative expected second derivatives of the log-likelihood function, considering N respondents, of the discrete choice model to be estimated (see Train, 2009). By manipulating the attribute levels of the alternatives, for known (assumed) parameter values, the analyst is able to minimize the elements within the AVC matrix, which in the case of the diagonals means lower standard errors and hence greater reliability in the estimates at a fixed sample size (or even at a reduced sample size).

In taking this approach, these authors have remained consistent with the underlying theory of experimental design as defined previously. Indeed the theory for generating SC experimental designs has as its objective the same objective when dealing with linear models; that is the minimizing of the variances and covariances of the parameter estimates. What is different, however, are the econometric models to which the theory is being applied. As discussed above, other differences have also emerged related to various assumptions that are required to be made when dealing with data specifically generated for logit type models. How various researchers have dealt with these assumptions is discussed in the next section.

5 THE WHO, THE WHAT AND THE WHY

In this section, we provide an abridged historical overview of research into the experimental design theory as applied to SC-type data. It is acknowledged that there exists a vast number of papers and researchers who have examined this issue, making it impossible to discuss in detail all the developments. For example, a group of researchers consisting of Grossmann, Grasshoff, Holling and Schwabe, among others, have been extremely active since 2001 and continue to be so. Nevertheless, the design strategies explored by these researchers, while arguably superior to some of those discussed herein, are less well known outside of the mainstream statistical literature. Hence, while inclusion of the research undertaken by this group, and others, would greatly benefit in obtaining a full understanding of experimental design theory for SC data, inclusion of such material would be beyond the knowledge of most individuals working within this field. Further, examining such work as well as other notable omitted research streams would represent a separate paper in and of itself. As such, what are perceived to be the major contributions that are commonly known about within the wider discrete choice community are concentrated on. This, however, is not meant to detract from the research of those not discussed at length herein.

Post Louviere and Hensher (1983) and Louviere and Woodworth (1983) to 1988

As discussed in section 4, the first SC studies focused mainly on introducing the method and promoting its benefits over the standard stated preference techniques used at the time (such as traditional conjoint methods). These early studies, therefore, did not concern themselves specifically with experimental design issues and simply borrowed design construction methods from elsewhere. As it turned out, elsewhere happened to be from the very methods that SC methods sought to replace; traditional conjoint design methodology. Traditional conjoint studies involve respondents ranking or rating alternatives (rather than picking one) constructed from either a full factorial or fractional factorial design which are not grouped together in choice tasks but presented all at once and are estimated using linear models such as linear regression (multivariate analysis of variance – MANOVA – was also popular at one stage). As such, the experimental design theory at the time focused largely on linear regression type models used for this type of data. This is not to suggest that research into aspects associated with the specific use of orthogonal designs as applied to discrete choice data were not undertaken in the early years of SC studies. For example, Anderson and Wiley (1992) and Lazari and Anderson (1994) looked at orthogonal designs capable of addressing problems of availability of alternatives. See Louviere et al. (2000) for a review of orthogonal design theory as applied to SC methods.

The variance-covariance (VC) matrix of a linear regression model is given in equation (7.1).

$$VC = \sigma^2 (X'X)^{-1}, \tag{7.1}$$

where σ^2 is the model variance, and X is the matrix made up of attributes and other variables, including interaction terms to be used in estimation.

Fixing the model variance for the present (which simply acts as a scaling factor), the elements of the VC matrix for linear regression models will generally be minimized when the columns of the X matrix are orthogonal. As such, when such models are estimated, orthogonality of data is considered important as this property ensures that (a) the model will not suffer from multicollinearity, and (b) the variances (and covariances) of the parameter estimates are minimized. As such, orthogonal designs, at least in relation to linear models, meet the two criteria for a good design mentioned earlier. They allow for an independent determination of each attribute's contribution on the dependent variable and they maximize the power of the design to detect statistically significant relationships (that is, maximize the *t*-ratios at any given sample size). Of course, the role that σ plays may be important and, as such, cannot always be ignored as suggested above. This is because it may be possible to locate a non-orthogonal design which produces non-zero covariances and slightly larger variances, but have smaller elements overall when scaled by sigma (this is because sigma is the error variance of the model, which is not independent of the betas). Nevertheless, orthogonal designs will tend to perform well overall for this type of model.

Despite the fact that non-linear models are commonly used to analyse discrete choice data, the question as to whether designs generated for linear models might be appropriate for such data remained surprisingly uncommented upon for a number of years. Where an examination of the problem was made, often inappropriate analysis was conducted that resulted in the not surprising conclusion that orthogonal designs are preferred to non-orthogonal designs. For example, Kuhfeld et al. (1994) compared attribute level balanced and unbalanced orthogonal designs with non-orthogonal designs using the information matrix associated with linear models (specifically equation (7.1) above without σ) despite applying the designs to (non-linear) logit models. It is little surprising that while they concluded that 'preserving orthogonality at all costs can lead to decreased efficiency', particularly when a balanced orthogonal design was not available, that 'non-orthogonal designs will never be more efficient than balanced orthogonal designs, when they exist' (Kuhfeld et al., 1994, p. 548).

Such misconceptions continue to this day. To demonstrate, consider the frequent practice of either (i) reporting the following design statistic in SC studies or (ii) the use of the statistic itself as the objective function to be maximized when generating a SC design (for example, Kuhfeld et al., 1994; Lusk and Norwood, 2005)

$$D\text{-efficiency} = \frac{100}{S \cdot \det(X'X)^{-1/K}}, \tag{7.2}$$

where S is the number of observations (that is, choice sets), K is the number of parameters in the design, X the design matrix, and det(.) refers to the determinant operator.

This measure is uninformative with respect to the operating conditions of discrete choice modelling under random utility theory because equation (7.2) is derived under the assumption that the model to be estimated is linear in the parameters. The relationship between this equation and that of the VC matrix of the homoscedastic linear regression model, $\sigma^2(X'X)^{-1}$ clearly demonstrates the relationship between the two. Indeed, equation (7.2) will return a value of 100 per cent for an orthogonal design and lower values for non-orthogonal designs. However, design orthogonality of this type does not necessarily imply statistical efficiency of discrete choice models, which are intrinsically non-linear in the parameters.

Nevertheless, it is important to note that the apparent success (or rather a lack of failure) of studies applying such designs meant that the use of orthogonal designs remained mostly unchallenged. Even with increasing evidence that non-orthogonal designs might be more appropriate for discrete choice models, orthogonal designs remain largely entrenched within the literature and continue to be the most common design method used.

Fowkes, Toner and Wardman et al. (1988–2000)

In the late 1980s, writing in the transport literature, researchers from the University of Leeds began to question the appropriateness of using orthogonal designs for discrete choice type data. In a series of research papers spanning over a decade, Fowkes, Toner and Wardman (among others) question the use of fractional factorial designs based on orthogonal arrays and discuss the importance of experiments that are realistic and make sense to the respondents as well as improving the robustness of the parameter estimates. Dealing specifically with binary choice tasks, the Leeds group designs were generated under the assumptions of non-zero priors for both generic and alternative-specific parameters. Such designs are referred to as locally optimal as the parameter priors are assumed to be known with certainty and the designs are optimized precisely for these parameter values. If the true parameter estimates differ to the assumed parameters (that is, the parameter priors were misspecified), then the design will lose efficiency. Rather than concern themselves directly with minimizing the standard errors of the parameter estimates, the Leeds group designs sought to minimize the variances of ratios of parameters (that is, they are concerned with willingness to pay – WTP – issues) which they were able to calculate from the model AVC matrix using the Delta-method.

It is worth noting, that even if a prior parameter is perfectly specified, the design need not be optimal. The most likely reasons for this to occur are due to constraints placed on the design by the user which may prevent the design from being truly optimal. Even if a design is optimal within the bounds of any such constraints, it is possible that smaller standard errors for one or more parameters might be possible if a different prior parameter is chosen, due to the non-linearity of the model. Likewise, misspecification of the prior will not necessarily imply that the efficiency of the design will be zero. This is because the efficiency of a design is dependent upon all parameters, and failure to correctly specify one parameter does not necessarily render impossible the ability to estimate other effects. Indeed, only in the case of perfect multicollinearity (or very near so) could the efficiency of the design be considered zero as this would render impossible the ability to estimate one or more parameters of the design. Further, it is important to acknowledge that even if the parameter prior is misspecified, the efficiency of the design might improve as the true parameter estimate might result in the design producing better sets of choice probabilities (that is, the magic Ps discussed later). That is, given analyst imposed constraints, such as attribute level balance, the elements of the AVC matrix may actually be smaller if a different set of priors were assumed or estimated than under the exact set of parameter priors used during the design generation process.

Rather than assume pre-defined fixed attribute levels, this class of designs were generated so as to allow the attribute levels of the design to take any value, including non-integers, and hence be continuous. In letting go of specific predetermined attribute levels and allowing for the attributes to take any value, it is possible to locate a design that will optimize the objective function of interest using a number of mathematical techniques such a non-linear programming. The Leeds group was able to utilize such methods to locate designs that minimized the variance of the ratio of two parameters and as such, generate designs which can be considered to be optimal under the assumptions for which they were generated.

Careful examination of the designs that were generated by this group led to the observation that many of the resulting choice tasks were not realistic from the perspective of the respondent (that is, the particular combinations of levels were not deemed to be realistic, nor the precision of the values shown, for example, a price of $5.37). For this reason additional requirements were imposed on the generated designs in which a reasonable coverage of so-called 'boundary values' were sought and obtained (see Fowkes and Wardman, 1998; Fowkes et al., 1993; Toner et al., 1998, 1999; and Watson et al., 2000, for further discussion of these designs). Further examination of these designs by the Leeds group found that they tended to retrieve very specific choice probabilities, which they referred to as 'magic Ps'. This finding was later independently rediscovered by other researchers working in other discipline areas, in particular by Kanninen in 2002, and later expanded by Johnson et al. in 2006.

Bunch, Louviere and Anderson (1996)

In 1996, two similar papers appeared simultaneously in the marketing literature dealing with experimental design theory as related to SC data (Bunch et al., 1996 and Huber and Zwerina, 1996). The first we discuss here (Bunch et al., 1996) appears only in the form of a working paper. This paper dealt specifically with strategies for generating designs for multinomial logit (MNL) models assuming either zero or non-zero local priors for generic parameter estimates. Unlike the earlier work coming out of the Leeds group, Bunch et al. assumed fixed attribute levels using orthogonal polynomial coding in estimation. It is worth noting that the specific coding structure can impact upon the efficiency of a design in a number of ways. First, non-linear coding will generally require more parameters and, hence, not only larger designs, but also will produce different AVC matrices than if linear codes are assumed. Secondly, the coding structure may influence the choice probabilities if the parameter priors are not adjusted to account for differences in the values taken over the design. Independent of the coding structure, the objective function used by Bunch et al. involved minimizing the elements of the resulting AVC matrix rather than with the variances of ratios of parameter estimates.

Bunch et al. (1996) promoted the use of the *D*-error statistic applying it to the expected

AVC matrix of the design as constructed for the MNL model. The *D*-error statistic, not to be confused with the *D*-efficiency measure (equation 7.2) suggested by Kuhfeld et al. (1994), is calculated by taking the determinant of the AVC matrix assuming a single respondent, Ω_1, and normalizing this value by the number of parameters, K. Minimizing the *D*-error statistic corresponds to minimizing, on average, the elements contained within the expected AVC matrix. Designs which minimize the *D*-error statistic are therefore called *D*-optimal designs. The *D*-error statistic is shown as equation (7.3).

$$D\text{-error} = \det(\Omega_1)^{1/K}. \tag{7.3}$$

Keeping in line with earlier empirical work in SC, Bunch et al. (1996) searched only among orthogonal designs. In doing so, they considered both simultaneously and sequentially constructed orthogonal designs in the generation process. A simultaneous orthogonal design is one where the attributes of the design are orthogonal not only within alternatives, but also between. This requires that the design be generated simultaneously for all alternatives. A sequentially constructed orthogonal design is one where the attributes of the design may be orthogonal within an alternative, but not necessarily between alternatives (see Louviere et al., 2000). As such, Bunch et al. (1996) kept the same properties as orthogonal designs, including attribute level balance constraints.

Unlike the designs generated by the Leeds group, the use of pre-specified fixed attribute levels makes it generally difficult to locate the design matrix which will be optimal. As such, algorithms are required which search over the design space by re-arranging the attribute levels of the design and testing the efficiency measure after each change. Only if all possible designs are tested can one conclude that the design is optimal. For designs with large design dimensions, this is often not possible and for this reason, such designs are more correctly referred to as efficient designs instead of optimal designs. Given that Bunch et al. (1996) considered designs which are orthogonal, only a subset of all possible designs were examined. For this reason the class of designs generated by Bunch et al. is more correctly referred to as locally optimal orthogonal *D*-efficient designs, as opposed to *D*-optimal designs. Although algorithms for locating SC designs are important, and formed a central part of the Bunch et al. paper, for reasons of space we do not discuss this aspect of the design generation process here (see Kessels et al., 2006 for an excellent discussion of design algorithms).

Huber and Zwerina (1996)

At the same time as Bunch et al. (1996), a paper by Huber and Zwerina (1996) appeared in the marketing literature. This paper covered much of the same material discussed in Bunch et al. (1996), however, a number of important and often subtle differences do exist between the two papers. Before discussing these differences, it is worth noting the many similarities between the two papers. As with the work of Bunch et al., Huber and Zwerina concerned themselves with optimal designs specifically generated for the MNL model assuming non-zero local priors for generic parameter estimates although they assumed effects coded variables as opposed to orthogonal polynomial coding. Further, as with Bunch et al., they assumed fixed attribute levels drawn from the underlying experimental design. Finally, they also concluded that minimizing the *D*-error statistic provided the best designs (in terms of generating statistically efficient designs).

Option A			Option B		
A1	**A2**	**A3**	**A1**	**A2**	**A3**
10	3	2	10	3	2
20	−5	4	30	−5	6
30	3	6	20	−5	4
10	−5	6	20	3	4
30	−5	2	10	−5	2
20	3	4	30	3	6

(a)

Option A			Option B		
A1	**A2**	**A3**	**A1**	**A2**	**A3**
10	3	2	10	3	2
20	−5	4	20	−5	4
30	3	6	30	3	4
30	−5	6	10	3	4
30	−5	2	10	−5	2
20	−5	6	20	3	6

(b)

Figure 7.1 Different definitions of attribute level balance

It is the differences between the two papers and the reaction of the literature to these differences, however, which is more telling. Unlike Bunch et al. (1996), Huber and Zwerina (1996) did not confine the design space to consist only of orthogonal designs. In letting go of orthogonality as a design criterion (or more precisely a constraint), they also relaxed the concept of attribute level balance. Whereas previously attribute level balance assumed that each level appeared an equal number of times within every column of the design (for a strict definition of attribute level balance, see Figure 7.1(a)), a new definition of attribute level balance was adopted that assumes that the level will appear an equal number of times for an attribute, independent of which alternative that attribute appears in (and hence levels must appear an equal number of times across columns of the design but not necessarily within each column; see Figure 7.1(b)). To demonstrate, consider the design in panel (a), in which the levels 10, 20 and 30 each appear twice for attribute A1 for both options A and B. Likewise, the levels of A2 and A3 appear an equal number of times in each column, independent of the alternative to which they belong. In the second design shown in panel (b), the levels do not appear an equal number of times within each column, however over both options A and B, each level appears exactly four times each within the corresponding attributes.

Two important and far reaching findings were obtained from this paper, however. First, Huber and Zwerina found that under non-zero local priors, non-orthogonal designs produced better designs in the form of statistically more efficient outcomes. While important, it is the second finding which has had a larger, if less desirable, impact on the literature. In their paper, Huber and Zwerina concluded that designs that produce roughly equal choice probabilities among the J alternatives were more statistically efficient than designs that resulted in less equal distributions of the choice probability. This finding unfortunately contradicts the earlier findings obtained from the Leeds group as well as later work of other researchers in this field who found that under certain conditions (actually the same conditions assumed by Huber and Zwerina), optimal designs are obtained not when the choice shares of the alternatives are equal to $1/J$, but rather optimal designs result from constructing alternatives that will produce certain choice probabilities which are not probability (or utility) balanced (the magic Ps discovered by the Leeds group). Unfortunately, it was the message of utility or probability balance which gained wider traction within the literature and not that letting go of orthogonality may result in statistical gains when applied to SC models. To this day, a number of papers continue to generate probability balanced designs when assuming non-zero priors, which will typically lead to inefficient designs.

Sándor and Wedel (2001, 2002, 2005)

A number of years passed before the next significant breakthrough occurred within the literature, again within marketing. In 2001, Sándor and Wedel introduced Bayesian efficient designs to the SC design field. Assuming an MNL model, with generic parameters applied to effects coded variables and fixed attribute levels, Sándor and Wedel (2001) relaxed the assumption of perfect a priori knowledge of the parameter priors through adopting a Bayesian like approach to the design generation process. Rather than assuming a single fixed value for each parameter prior, the efficiency of the design is calculated over a number of simulated draws taken from prior parameter distributions assumed by the analyst. Different distributions may be associated with different population moments representing different levels of uncertainty with regards to the true parameter values. In this way, by optimizing the efficiency of the design over a range of possible parameter prior values (drawn from the assumed parameter prior distributions) the design is made more robust, at least within the range of the assumed distributions. Such designs will be generally less efficient than an equivalent locally optimal design (represented by the solid line) but will be more robust to prior parameter misspecification. As with Huber and Zwerina (1996), they found non-orthogonal designs outperformed orthogonal designs based on the Bayesian equivalent of the *D*-error statistic.

In subsequent research Sándor and Wedel (2002, 2005), assuming generic parameter estimates for effects coded variables and fixed attribute levels as well as local (fixed) prior parameter estimates, derived the AVC matrix for the cross-sectional version of the mixed MNL (MMNL) model. As such, they were the first to generate designs for a model other than the MNL model. In doing so, they retained the use of the *D*-error statistic as their design criterion and, as such, despite differences in assumptions made in terms of model type and how the prior parameters are generated, retained as their design objective, the desire to locate a design that results in smaller standard errors (and covariances).

Street and Burgess (2001 to Current)

An independent stream of research on generating designs for SC studies began to appear in 2001 within the statistics and marketing literatures and centers around what can be referred to as Street and Burgess type designs (see Burgess and Street, 2005; Street and Burgess, 2004, 2007; Street et al., 2001, 2005). Like earlier researchers, Street and Burgess assume an MNL model specification in deriving the AVC matrix for their designs; however, the mathematical derivations used to obtain to the AVC matrix are performed in a somewhat different manner. Whereas other researchers derive the second derivatives with respect to the parameter β such that

$$\Omega = I^{-1}, \text{ with } I = -E\left(\frac{\partial^2 \log L}{\partial\beta\,\partial\beta'}\right), \tag{7.4}$$

where $E(\cdot)$ is used to express the large sample population mean, Street and Burgess calculate the second derivatives with respect to total utility V, such that

$$\Omega = I^{-1}, \text{ with } I = -E\left(\frac{\partial^2 \log L}{\partial V \partial V'}\right), \tag{7.5}$$

This difference in the mathematical derivations of the AVC matrix has resulted in significant confusion within the literature, with claims that the Street and Burgess approach is unrelated to the more mainstream SC experimental design literature as discussed herein. This view has been further enhanced given the fact that the resulting matrix algebra used to generate the AVC matrices under the two derivations appear to be very different. However, Bliemer and Rose (2010b) were able to show that Street and Burgess designs are a special case of the more general methods used by other researchers, as described earlier.

Aside from the assumption of an MNL model specification, Bliemer and Rose (2010b) were able to reproduce Street and Burgess type designs using the same methods used by other researchers discussed above, if they assumed the data were coded using an orthonormal coding structure. Beginning with a sequentially generated orthogonal design using the methods first described by Bunch et al. (1996), Street and Burgess designs can be constructed after first converting the design (see Table 7.1a) into orthogonal contrast codes (Table 7.1b) (see Keppel and Wickens, 2004). The orthogonal contrast codes are then converted into the orthonormal coding structure by first computing the sum of the squares of each column (shown at the base of Table 7.1b) and next dividing each column of the orthogonal contrast code by this number (see Table 7.1c).

The AVC matrix of the design is then computed under the assumption of zero local priors, assuming that the parameters are generic across alternatives. The elements of the resulting AVC matrix are then normalized by dividing each value by the product of the number of levels, L_k, of each attribute k of the original design (that is, by $\prod_{k=1}^{K} L_k$). The design can then be optimized using the same D-error measure promoted by other researchers.

The major contribution of Street and Burgess, however, has been to derive a method that can be used to locate the optimal design under the above set of assumptions without having to resort to complex (iterative) algorithms. They showed that by using so-called design generators, which are sequences of numbers, to create the attribute levels for the second alternative (and third, and so on) they ensure that attribute levels will have minimum overlap (that is, are mostly different for each alternative). They show that the best (minimum) D-error can be determined, such that the D-efficiency of a design can be computed relative to this minimum D-error. Then, let I be the Fisher information matrix of a certain design (assuming some design generator), then D-efficiency can be computed as

$$D\text{-efficiency} = \frac{det(I)}{det(I_{\max})} \times 100. \tag{7.6}$$

Equation (7.6) provides a measure, as a percentage, as to how efficient a design is under the specific assumptions outlined above; that is the design is constructed assuming

Table 7.1 Design codes to orthogonal contrast codes

(a) Design codes

S	A1	A2	B1	B2
1	0	0	2	1
2	1	1	0	0
3	2	1	1	0
4	2	0	1	1
5	0	0	2	1
6	1	1	0	0
7	1	0	0	1
8	2	0	1	1
9	0	1	2	0
10	1	0	0	1
11	2	1	1	0
12	0	1	2	0

(b) Orthogonal contrast codes

S	A1a	A1b	A2a	B1a	B1b	B2a
1	−1	1	−1	1	1	1
2	0	−2	1	−1	1	−1
3	1	1	1	0	−2	−1
4	1	1	−1	0	−2	1
5	−1	1	−1	1	1	1
6	0	−2	1	−1	1	−1
7	0	−2	−1	−1	1	1
8	1	1	−1	0	−2	1
9	−1	1	1	1	1	−1
10	0	−2	−1	−1	1	1
11	1	1	1	0	−2	−1
12	−1	1	1	1	1	−1
	8	**24**	**12**	**8**	**24**	**12**

(c) Orthonormal coding

S	**A1a**	**A1b**	**A2a**	**B1a**	**B1b**	**B2a**
1	−0.35	0.20	−0.29	0.35	0.20	0.29
2	0.00	−0.41	0.29	−0.35	0.20	−0.29
3	0.35	0.20	0.29	0.00	−0.41	−0.29
4	0.35	0.20	−0.29	0.00	−0.41	0.29
5	−0.35	0.20	−0.29	0.35	0.20	0.29
6	0.00	−0.41	0.29	−0.35	0.20	−0.29
7	0.00	−0.41	−0.29	−0.35	0.20	0.29
8	0.35	0.20	−0.29	0.00	−0.41	0.29
9	−0.35	0.20	0.29	0.35	0.20	−0.29
10	0.00	−0.41	−0.29	−0.35	0.20	0.29
11	0.35	0.20	0.29	0.00	−0.41	−0.29
12	−0.35	0.20	0.29	0.35	0.20	−0.29

an MNL model specification, using orthonormal coding with local priors for generic parameters equal to zero. Unfortunately, in a case of selectively choosing criteria to promote one design paradigm over another, the measure has been incorrectly applied by some to infer that designs generated under different sets of assumptions are not optimal. That is, these equations should only be used to optimize designs under the specific conditions that the equations were derived for, and not to infer anything about designs generated under other sets of assumptions.

Kanninen (2002, 2006)

In 2002, Kanninen independently rediscovered the fact that *D*-optimal designs generated under the assumption of an MNL model specification with non-zero local priors for generic parameter estimates tend to retrieve specific non balanced choice probabilities. As with the work of the Leeds group, Kanninen was able to show analytically that utility or probability balance in choice tasks represent an undesirable property, and in doing so suggest rules that minimize the variance of estimates in an optimal manner, based on desirable choice probabilities or what the Leeds group referred to as magic Ps. Kanninen (2002) proposed a design approach where the first $K - 1$ attribute levels are first generated for each of the alternatives, typically using an orthogonal design or Street and Burgess design. The level of the last K^{th} attribute for each alternative is then generated as a continuous variable (usually a price attribute). The values of these continuous variables are chosen such that the choice probabilities take certain values that minimize the elements of the AVC matrix under the assumption of non-zero prior parameters.

Although the boundary value method of the Leeds group is somewhat different in derivation, the implications remain the same, and similar results are achieved. The main differences between the two methods, however, lie in the fact that the designs generated by the Leeds group were constructed so as to minimize the variances of the ratios of two parameters (that is, dealing with WTP issues), whereas the approach adopted by Kanninen works directly with the variances of the parameter estimates using the *D*-error measure. Another difference between the two approaches is that the Leeds group promoted allowing all attributes of the design to be treated as continuous (given transport applications where the attributes often considered are times and costs, this is possible), while Kanninen designs tend to allow only one attribute to be treated as continuous (again, this is in line with the literature where these designs have been applied – marketing and environmental economics – in which many of the attributes are qualitative in nature and it makes little to no sense to allow them to be shown to respondents as a continuous attribute level). Kanninen (2002) and Johnson et al. (2006) have determined the desirable probabilities for a limited number of designs with two alternatives (see Table 7.2), however, unlike the Leeds group, Kanninen's earlier work was not constrained to designs with only two alternatives.

One concern with these designs, however, is that they are optimal for the parameter priors assumed. While this is true of other locally *D*-optimal designs, the values that the continuous variables take are particularly sensitive in terms of the priors assumed,

Table 7.2 Optimal choice probability values for specific designs with two alternatives

Number of attributes (K)	Number of choice sets	Optimal choice-percentage split for two-alternative model
2	2	0.82 / 0.18
3	4	0.77 / 0.23
4	4	0.74 / 0.26
5	8	0.72 / 0.28
6	8	0.70 / 0.30
7	8	0.68 / 0.32
8	8	0.67 / 0.33

Source: Adapted from Johnson et al. (2006).

and may change markedly given different parameter priors. Thus, Kanninen (2002) recommends a continual process of updating the design once data is collected and more likely parameter estimates are obtained. Further, it is worth stressing that the results of Kanninen (and the Leeds group) hold only for designs assuming generic parameter estimates. Once alternative specific parameter priors are assumed, it is no longer possible to calculate desirable choice probabilities even for the MNL model, as these will differ over each choice task.

Bliemer, Rose and Scarpa (2004 to Current)

Writing in the transportation and environmental economics literature, Bliemer, Rose and Scarpa have sought to extend the theory of experimental design to more advanced discrete choice models as well as address issues related to the sample size requirements for these types of studies. In 2004, Rose and Bliemer began by examining the impact that relaxing the assumption of orthogonality has on the performance of logit models; in particular MNL models assuming non-zero local priors (Rose and Bliemer, 2004). In line with earlier work, they concluded that orthogonality as a design principle did not appear to be a desirable property for the non-linear logit models. Bliemer and Rose (2005a) and Rose and Bliemer (2005a) next sought to extend the methods advocated by Bunch et al. (1996) for MNL model specifications assuming non-zero local priors to allow for alternative specific and generic parameter estimates. Concurrently, Bliemer and Rose (2005b) and Rose and Bliemer (2005b) turned their attention towards issues of sample size requirements for SC experiments. Throughout this work, fixed attribute levels were assumed in the design generation process.

Bliemer and Rose pointed out that the AVC matrix for discrete choice models is inversely related to the number of times a design is replicated within the data, that is, assuming all choice tasks are answered by each respondent, which is equivalent to the number of respondents N.[1] As such, the analyst can calculate the values contained within the AVC matrix for any sample size, simply by determining the AVC matrix for a single respondent and then dividing the resulting matrix by N. This means that the standard errors decrease with a rate of $1/\sqrt{N}$. Using this relationship, Figure 7.2(a) reveals the consequences of investing in larger sample sizes for a given design X^{I}. While initial gains can be achieved in terms of improvements to the expected asymptotic standard errors achieved from models estimated based on the design from adding more respondents, such improvements occur at a diminishing rate until each additional respondent added will have only a marginal impact on the expected asymptotic standard errors. Hence, increasing the sample size beyond a certain limit will typically have little impact upon the statistical significance of the parameter estimates achieved from SC studies. Figure 7.2(b) reveals the impact for a given set of population parameters of investing in a better design X^{II} (that is, more efficient design). Typically, larger decreases in the standard error can be achieved by investing in finding a more efficient design than by investing in a larger sample.

Given this, Bliemer and Rose were able to use this relationship to provide insight into the sample size requirements for SC experiments. Seeing that the square roots of the diagonal elements of the AVC matrix represent the asymptotic standard errors for the parameter estimates, and the asymptotic t-ratios are simply the parameter estimates

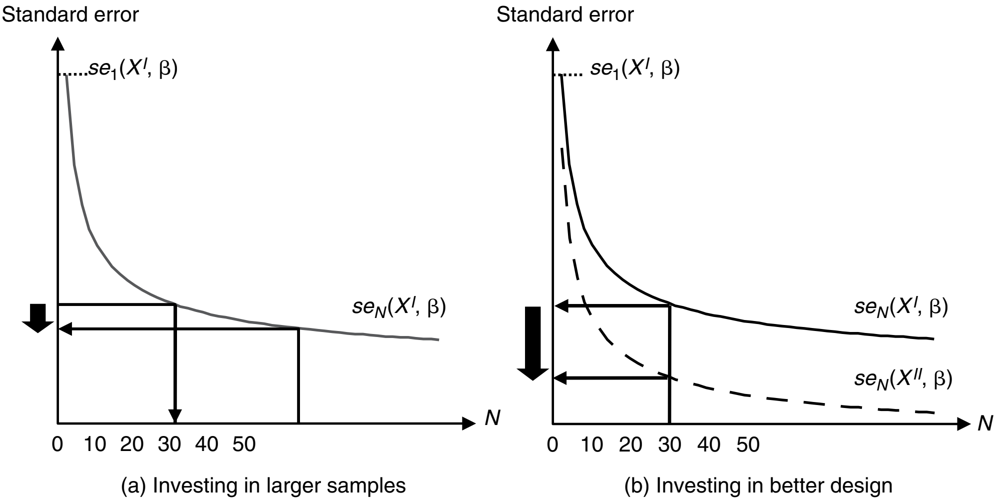

Figure 7.2 Comparison of investing in larger sample sizes versus more efficient designs

divided by the asymptotic standard errors (equation (7.7)), it is possible to determine the likely asymptotic t-ratios for a design assuming a set of prior parameter estimates.

$$t_k = \frac{\hat{\beta}_k}{se_k/\sqrt{N_k}}. \tag{7.7}$$

Rearranging equation (7.7),

$$N_k = \frac{t_k^2 se_k^2}{\beta_k^2}. \tag{7.8}$$

Equation (7.8) allows for a determination of the sample size required for each parameter to achieve a minimum asymptotic t-ratio, assuming a set of non-zero prior parameter values. To use these equations, the analyst might use the prior parameters used in generating the design, or test the sample size requirements under various prior parameter misspecifications. Once the sample size is determined for all parameters, the analyst can then select the sample size that will be expected to result in all asymptotic t-ratios taking a minimum pre-specified value (for example, 1.96). Such designs are called S-efficient designs. Bliemer and Rose noted, however, that sample sizes calculated using this method should be considered as an absolute theoretical minimum. The method assumes certain asymptotic properties that may not hold in small samples. Further, the method does not consider the stability of the parameter estimates, nor at what sample size parameter stability is likely to be achieved (that is, how many respondents are required before the parameter estimates reflect the sample population estimates). Comparing samples sizes using equation (7.8) for different parameters may also give an indication of which parameters will be more difficult to estimate (at a certain level of significance) than other parameters.

Rose and Bliemer (2006) extended the theory of SC designs next to include covariates (such as socio-demographics; Jaeger and Rose 2008 later looked at including contextual covariates into the design, such as situational factors usually held constant within the survey preamble) in the utility functions and hence also in the AVC matrices of the designs. Assuming an MNL model with non-zero local priors and combinations of alternative specific and generic parameters, they were able to demonstrate a method capable of jointly minimizing the elements of the AVC matrix while determining the optimal number of respondents to sample from different segments. This was accomplished by determining optimal weights to apply to different segments of the Fisher information matrix based on how many respondents belong to each segment.

Rose et al. (2008) next looked at SC studies requiring pivot (or customized) designs where the levels of the design alternatives are computed as percentage differences from some pre-specified respondent specific status quo alternative, rather than as specific predefined levels chosen by the analyst. Again, assuming an MNL model specification with non-zero local priors and combinations of alternative specific and generic parameters, they explored a number of design procedures capable of optimizing designs at the individual respondent level.

Meanwhile, Ferrini and Scarpa (2007), writing in the environmental economics literature, extended the optimal design theory to the panel version of the error component models assuming non-zero local priors and fixed attribute levels. In considering the panel version of the model, Ferrini and Scarpa's paper represents a significant leap forward in the theory of SC experimental design, as it was the first to consider the issue of within respondent preference correlations which theoretically exist over repeated choice tasks which cannot be detected, and indeed are ignored in models such as the MNL model. Unlike earlier papers, however, Ferrini and Scarpa used simulation to derive the AVC matrix of the model, rather than more common analytical derivations.

Scarpa and Rose (2008) looked at various design strategies assuming an MNL model specification with non-zero local priors and generic parameters. Unaware of the earlier work of the Leeds group, they also used the delta method to derive the variance of the ratio of two parameters and advocated optimizing on this measure if WTP was the primary concern of the study.

Bliemer et al. (2009) next applied experimental design theories assuming non-zero local priors and Bayesian priors with designs assuming fixed attribute levels in order to estimate nested logit (NL) models. They were able to show that designs generated assuming an optimal design that has been derived for a MNL model does not necessarily perform well for a nested model with correlated error terms within the different nests of alternatives.

In 2010, Bliemer and Rose were able to analytically derive the AVC matrix for the panel version of the MMNL model. Via a number of case studies in which both non-zero local priors and Bayesian priors with both alternative specific and generic parameters, they were able to compare the design efficiencies for designs generated assuming MNL, cross sectional MMNL and panel MMNL model specifications. They found that efficient designs generated specifically for the MNL model are mostly also quite efficient for the panel MMNL model, and vice versa. However designs generated for a cross sectional MMNL specification were very different in terms of the resulting statistical efficiency and sample size requirements (see Bliemer and Rose, 2010a).

One criticism often levelled at those advocating the generation of the efficient design approach to SC studies is the need to know in advance the precise econometric model that will be estimated once data has been collected. There unfortunately exist many forms of possible discrete choice models that analysts may wish to estimate once SC data has been collected (for example, MNL, NL, GEV, MMNL). The log-likelihood function will differ for different model types and hence, given that the AVC matrix is mathematically given as the inverse of the second derivatives of the log-likelihood function, the AVC matrix for each type of model will also be different, even if the same design is used. As such, the construction of efficient designs requires not only an assumption as to the parameter priors assumed, but also what AVC matrix the analyst is attempting to optimize.

Given differences in the AVC matrices of different discrete choice models, attempts at minimizing the elements of the AVC matrix assuming one model, may not necessarily minimize the elements of the AVC matrix of another model. Similar to the problem involving knowledge of the parameter estimates with certainty (given by local priors), the analyst is unlikely to know precisely what model is likely to be estimated in advance. The problem then becomes one of having to select the most likely model that will be estimated once data has been collected. To address this specific issue, Rose et al. (2009) advocated the use of a model averaging approach, where different weights could be applied to the Fisher information matrices obtained assuming different model specifications given a common design. Included in the model averaging process were MNL, cross sectional error components and MMNL, and panel error components and MMNL model specifications.

Bliemer and Rose (2010c) looked at sequential designs in which the design is optimized after collecting data from a respondent, based on the current parameter estimates as priors, and given to the next respondent. Such adaptive designs are often said to be prone to endogeneity bias. By only adapting the design between respondents, however, and not within-respondent, this endogeneity bias can be mostly avoided. They found that starting with zero local priors and then sequentially generating more efficient designs based on the parameter estimates after *n* respondents, is a very efficient way of collecting data. However, as designs found at the early stage of data collection are sensitive to likely unstable priors, parameter biases could be found in small sample sizes, but disappear in larger sample sizes. Starting with an orthogonal design for the first, say, 20 respondents, and then generate sequentially optimal designs for the next respondents will avoid these biases in small sample sizes, while being able to minimize the standard errors significantly.

Most recently, Rose et al. (2011) sought to extend upon the earlier research originating from both the Leeds group and Kanninen to a wider range of SC problems. Unfortunately, they found that it was only possible to derive optimal choice probabilities for designs generated under the assumption of a MNL model specification and non-zero local priors parameters and generic parameter estimates. To overcome these limitations, they demonstrated how the Nelder–Mead algorithm could be used to locate the optimal choice probabilities for any model type with any number of alternatives and any type of prior parameters, including non-zero Bayesian priors. In contrast to the designs with generic parameters, fixed magic Ps do not seem to exist for this more general case.

Kessels, Goos, Vandebroek and Yu (2006 to Current)

Research groups centred at the University of Antwerp and Catholic University of Leuven have also been active in promoting the application of experimental design theory in the generation of SC experiments. A significant proportion of the work originating from this group deals specifically with algorithms for generating these types of designs and are hence beyond the scope of this current paper (see, for example, Kessels et al., 2006, 2009; Yu et al., 2010, 2012). This group has actively examined a wide range of design criterion beyond the *D*-error statistic, including *G*- and *V*-error measures which are designed to minimize the prediction error variance (see, for example, Kessels et al., 2006). However, this group has also been actively looking at other areas of SC experimental design theory. In particular, they have examined designs for the cross sectional MMNL model under non-zero Bayesian priors and generic parameters (see Yu et al., 2009) as well as designs generated for MNL models with no choice alternatives under the assumption of non-zero Bayesian priors (Vermeulen et al., 2008) and NL models including no choice alternatives under the assumption of non-zero Bayesian priors (see Goos et al., 2010).

Yu et al. (2008) also looked at efficient designs assuming non-zero priors for both main and interaction effects. This was later extended to allow for interaction effects assuming an MNL model specification and non-zero Bayesian priors. More recently, theory related to designs optimized for WTP to include non-zero Bayesian priors has been examined (Vermeulen et al., 2011a). This group has also been actively researching sequentially generated designs which adapt at the individual level both for models optimized for measuring preferences (Yu et al., 2011) and WTP (Danthurebandara et al., 2011a). Two new areas of innovation recently researched by this group include the inclusion of an entropy measure in generating SC designs for MNL models with non-zero priors as a measure of complexity (Danthurebandara et al., 2011b) and designs for rank order experiments involving rank exploded data (Vermeulen et al., 2011b).

6 CONCLUDING COMMENTS

The chapter has sought to provide a brief overview of experimental design theory as related to SC studies and, in doing so, make clear that the construction of experimental designs for SC problems is not independent of other wider decisions that are necessary when constructing these types of questionnaires. The central thesis of the chapter, however, has been to discuss how key research streams, as emanating from different research groups, have progressed knowledge in this field. The chapter has not sought to provide any new information on the design of SC experiments as everything discussed herein has already been reported elsewhere. However, in examining research from several different disciplines we have attempted to bring together, in one document, the substantive frontier of knowledge as held at this time.

In writing the chapter we have not attempted to provide an in-depth discussion or analysis of the specific methods used for generating SC designs. By reviewing the key research streams, references have been provided that the interested reader can use to learn the specifics of the various methods. Nevertheless, where necessary, we have pointed out key misunderstandings of certain research streams and attempted to resolve these within a general theory of design. For example, we have argued that Street and Burgess sought specifically to optimize designs under specific assumptions which can be replicated using the same methods advocated by other researchers and are a specific case of optimal designs rather than representing designs which will be optimal independent of the assumptions made.

One important point we hope the reader does take away from this chapter is the inappropriate use of a number of statistical measures which have come to be prevalent within the literature. In particular, we deliberately point out two such measures; the first designed specifically to optimize designs for linear models and the second used to optimize designs assuming an MNL model specification with all generic parameters and priors equal to zero under orthonormal codes (that is, the specific case examined by Street and Burgess). While use of these measures is perfectly valid if one wishes to optimize a design under these expressed assumptions, applying these measures to determine how optimal a design is generated under other assumptions (including different model specification, prior parameter assumptions and coding structures) is both incorrect and misleading (we are not implying that those who devised these measures have applied the measures incorrectly; however, we can attest to the fact that a number of reviewers have over time applied them inappropriately to infer designs are not optimal, even when generated under different sets of assumptions). To apply a sporting analogy to this situation to highlight how inappropriate the blanket use of such measures is, this is equivalent to attempting to play golf applying the rules of hockey simply because both sports utilize sticks.

In this chapter, it has been argued that the use of orthogonal designs for non-linear models, such as the logit model, will be inefficient under most, but not all, assumptions made during the design generation phase (for example, the specific case examined by Street and Burgess has shown that orthogonal designs are optimal under some assumptions). Nevertheless, orthogonal designs remain to this day the most widely used design type. Such prevalence is the result of the fact that orthogonal designs appear to (and actually do) work well in most cases and it is important to understand why this is the case.

Designs of all types, whether orthogonal or non-orthogonal, are generated under assumptions about the true population parameter estimates (that is, the priors that are assumed). These assumptions are either explicitly acknowledged by those generating the design or implicitly made without their knowledge. Perhaps unknown too many, an orthogonal design will be the optimal design for an unlabelled experiment under the assumption of local priors set to zero (see Bliemer and Rose, 2005a). If the true population parameters differ from those that are assumed in the design generation phase, then the design will generally lose statistical efficiency. If a prior parameter is incorrectly specified, typically an increase in the standard error obtained will occur for the true population parameter value, all else being equal. Note that this does not mean that the true parameter cannot be estimated by the design, but simply that a larger sample size would be required to detect statistical significance of the parameter estimate than otherwise would have been the case had the prior parameter assumed been correct. It is this precise reason that orthogonal designs have appeared to work well in the past and will likely continue to work well into the future. That is, the sample sizes used in practice have in most cases reported in the literature been such that they have sufficiently compensated the loss of efficiency in the design as the true parameters diverge from those assumed in generating the design. Nevertheless, as econometric models become more sophisticated, such an argument may become less valid and larger samples may be required to detect particular effects within the data. This may be particularly true of models using random parameters to detect preference or scale heterogeneity or models using complex non-linear functions. Still, the point of those advocating non-orthogonal designs generated under non-zero prior parameter estimates is that in undertaking SC experiments, one would assume that the attributes chosen will have some influence in the choices made by the respondents, and hence the true population parameters will be non-zero. In such cases, the argument is that these designs will outperform orthogonal designs given similar sample sizes, or produce the same results as an orthogonal design but with smaller sample sizes. If parameters are unknown, but the signs known, then one can find more efficient designs than an orthogonal design.

It is important to note that the above discussion is predicated on the assumption of all else being equal. That is, it assumes that there exists no link between the population parameter estimates and the design itself. Several articles have convincingly argued that the design may result in unintended biases of the parameter estimates (for example, Louviere and Lancsar, 2009). In theory, however, this should not be the case. McFadden (1974) showed that asymptotically, the parameter estimates should converge to the population parameters, independent of the data matrix (that is, design in this instance). Using Monte Carlo simulations, McFadden further showed that this was the case in quite small finite samples, with as few as 50 choice observations. Numerous studies using simulation have led to the same conclusions (for example, see Ferrini and Scarpa, 2007). However, the arguments put forward by Louviere and Lancsar (2009), drawing on research derived from the psychology literature on demand characteristics or demand induced effects (see Orne, 1959, 1969), remain compelling. They posit that if the design attributes correlate with unobserved omitted covariates or latent contrasts, such as personality profiles or other such characteristics, then the resulting parameters obtained from different designs will indeed be influenced by the specific design used. Such biases will not exist in simulated data unless they are assumed in the data generation process, which makes empirical studies far more important to determining if these biases are real or not. Thus, this represents an important area of research that is urgently required as the existence of any such biases may require a different line of enquiry in terms of generating designs than has occurred in the past, as outlined in this chapter.

Similarly, the impact of designs upon scale also represents an important research area. Louviere et al. (2008) and Bliemer and Rose (2011) found scale differences across various designs relating to how ‘easy’ the resulting questions are as generated from the design. Both Louviere et al. and Bliemer and Rose found, for example, that orthogonal designs tended to lead to lower error variances than efficient designs possibly as a result of the presence of dominated alternatives. Given that efficient designs are less likely to have dominated alternatives than orthogonal designs, as choice probabilities are optimized, the questions arising from the use of orthogonal designs will be easier to answer, resulting in lower error variance. As such, there exists the very real possibility that any move away from orthogonal designs to other designs represents a trade-off between capturing more information per question versus lowering error variance. Once more, further research is required to address this specific issue.

ACKNOWLEDGEMENTS

We would like to thank Matthew Beck and Kirsten Howard for their comments on earlier drafts of this chapter. We would also like to acknowledge the anonymous reviewer who provided useful comments and feedback resulting in a much better chapter. All mistakes and errors remain ours however.

NOTE

1. If a design is blocked, then each design replication will require several respondents, and hence the total number of respondents will be $N \times b$, where b is the number of blocks.

REFERENCES

Anderson, D.A. and J.B. Wiley (1992) 'Efficient choice set designs for estimating cross-effects models', *Marketing Letters*, **3** (4), 357–70.

Bliemer, M.C.J., J.M. Rose and D.A. Hensher (2009) 'Constructing efficient stated choice experiments allowing for differences in error variances across subsets of alternatives', *Transportation Research Part B*, **43** (1), 19–35.

Bliemer, M.C.J. and J.M. Rose (2005a) 'Efficient designs for alternative specific choice experiments', working paper, ITS-WP-05-04, Institute of Transport Studies, University of Sydney.

Bliemer, M.C.J. and J.M. Rose (2005b) 'Efficiency and sample size requirements for stated choice studies', working paper, ITS-WP-05-08, Institute of Transport Studies, University of Sydney.

Bliemer, M.C.J. and J.M. Rose (2009) 'Designing stated choice experiments: the state of the art', in R. Kitamura, T. Yoshi and T. Yamamoto (eds), *The Expanding Sphere of Travel Behaviour Research, Selected Papers from the 11th International Conference on Travel Behaviour Research*, Bingley: Emerald, pp. 495–8.

Bliemer, M.C.J. and J.M. Rose (2010a) 'Construction of experimental designs for mixed logit models allowing for correlation across choice observations', *Transportation Research Part B*, **46** (3) special issue, 720–34.

Bliemer, M.C.J. and J.M. Rose (2010b) 'A unified theory for stated choice designs', Institute of Transport and Logistics Studies working paper series, University of Sydney.

Bliemer, M.C.J. and J.M. Rose (2010c) 'Serial choice conjoint analysis for estimating discrete choice models', in S. Hess and A. Daly (eds), *Choice Modelling: The State-of-the-Art and the State-of-the-Practice*, Bingley: Emerald, pp. 139–61.

Bliemer, M.C.J. and J.M. Rose (2011), 'Experimental design influences on stated choice outputs: an empirical study in air travel choice', *Transportation Research Part A*, **45** (1), 63–79.

Bunch, D.S., J.J. Louviere and D.A. Anderson (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, L. and D.J. Street (2005), 'Optimal designs for choice experiments with asymmetric attributes', *Journal of Statistical Planning and Inference*, **134** (1), 288–301.

Carlsson, F. and P. Martinsson (2003), 'Design techniques for stated preference methods in health economics', *Health Economics*, **12** (4), 281–94.

Danthurebandara, V., J. Yu and M. Vandebroek (2011a), 'Sequential choice designs to estimate the distribution of willingness-to-pay', *Quantitative Marketing and Economics*, **9**, (4), 429–48.

Danthurebandara, V., J. Yu and M. Vandebroek (2011b), 'Effect of choice complexity on design efficiency in conjoint choice experiments', *Journal of Statistical Planning and Inference*, **141** (7), 2276–86.

Ferrini, S. and R. Scarpa (2007), 'Designs with a priori information for nonmarket valuation with choice-experiments: a Monte Carlo study', *Journal of Environmental Economics and Management*, **53** (3), 342–63.

Flynn, T.N., J.J. Louviere, T.J. Peters and J. Coast (2008), 'Estimating preferences for a dermatology consultation using best-worst scaling: comparison of various methods of analysis', *BMC Medical Research Methodology*, **8** (76), 1–12.

Fowkes, A.S. and M.R. Wardman (1988), 'The design of stated preference travel choice experiments with particular regard to inter-personal taste variations', *Journal of Transport Economics and Policy*, **22** (1), 27–44.

Fowkes, A.S., M. Wardman and D.G.P. Holden (1993), 'Non-orthogonal stated preference design', *Proceedings of the PTRC Summer Annual Meeting*, London: PTRC Education Research Services Limited, pp. 91–7.
Goos, P., B. Vermeulen and M. Vandebroek (2010), 'D-optimal conjoint choice designs with no-choice options for a nested logit model', *Journal of Statistical Planning and Inference*, **140** (4), 851–61.
Hensher, D.A. and J.J. Louviere (1983), 'Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event', *Journal of Consumer Research*, **10** (3), 348–61.
Huber, J. and K. Zwerina (1996), 'The importance of utility balance and efficient choice designs', *Journal of Marketing Research*, **33** (3), 307–17.
Jaeger, S.R. and J.M. Rose (2008), 'Stated choice experimentation, contextual influences and food choice: a case study', *Food Quality and Preference*, **19** (6), 539–64.
Johnson, F.R., B.J. Kanninen and M. Bingham (2006), 'Experimental design for stated choice studies', in B.J. Kanninen (ed.), *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice*, Dordrecht: Springer, pp. 159–202.
Kanninen, B.J. (2002), 'Optimal design for multinomial choice experiments'. *Journal of Marketing Research*, **39** (2), 214–17.
Keppel, G. and D.W. Wickens (2004), *Design and Analysis: A Researchers Handbook*, 4th edn, Englewood Cliffs, NJ: Pearson Prentice Hall.
Kessels, R., B. Bradley, P. Goos and M. Vandebroek (2009), 'An efficient algorithm for constructing Bayesian optimal choice designs', *Journal of Business and Economic Statistics*, **27** (2), 279–91.
Kessels, R., P. Goos and M. Vandebroek (2006), 'A comparison of criteria to design efficient choice experiments', *Journal of Marketing Research*, **43** (3), 409–19.
Kuhfeld, W.F., R.D. Tobias and M. Garratt (1994), 'Efficient experimental design with marketing research applications', *Journal of Marketing Research*, **31** (4), 545–57.
Lazari, A.G. and D.A. Anderson (1994), 'Designs of discrete choice set experiments for estimating both attribute and availability cross effects', *Journal of Marketing Research*, **31** (3), 375–83.
Louviere, J.J. and D.A. Hensher (1983), 'Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event', *Journal of Consumer Research*, **10** (3), 348–61.
Louviere, J.J. and E. Lanscar (2009), 'Choice experiments in health: the good, the bad, and the ugly and toward a brighter future', *Health Economics, Policy and Law*, **4** (4), 527–46.
Louviere, J.J. and G. Woodworth (1983), 'Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data', *Journal of Marketing Research*, **20** (4), 350–67.
Louviere, J.J., D.A. Hensher and J.D. Swait (2000), *Stated Choice Methods: Analysis and Application*, Cambridge: Cambridge University Press.
Louviere, J.J., T. Islam, N. Wasi, D. Street and L. Burgess (2008), 'Designing discrete choice experiments: do optimal designs come at a price?', *Journal of Consumer Research*, **35** (2), 360–75.
Lusk, J.L. and F.B. Norwood (2005), 'Effect of experimental design on choice-based conjoint valuation estimates', *American Journal of Agricultural Economics*, **87** (3), 771–85.
MacCrimmon, K.R. and M. Toda (1969), 'The experimental determination of indifference curves', *Review of Economic Studies*, **36** (4), 433–51.
May, K.O. (1954), 'Intransitivity, utility, and the aggregation of preference patterns', *Econometrica*, **22** (1), 1–13.
McFadden, D. (1974), 'Conditional logit analysis of qualitative choice behaviour', in P. Zarembka (ed.), *Frontiers of Econometrics*, New York: Academic Press, pp. 105–42.
Mosteller, F. and P. Nogee (1951), 'An experimental measurement of utility', *Journal of Political Economy*, **59** (5), 371–404.
Orne, M.T. (1959), 'The demand characteristics of an experimental design and their implications', paper presented at the American Psychological Association, Cincinnati, September.
Orne, M.T. (1969), 'Demand characteristics and the concept of quasi-controls', in R. Rosenthal, R.L. Rosnow and A.E. Kazdin (eds), *Artifact in Behavioral Research*, New York: Academic Press, pp. 143–79.
Peirce, C.S. (1876), 'Note on the theory of the economy of research', *Coast Survey Report*, pp. 197–201.
Rose, J.M. and M.C.J. Bliemer (2004), 'The design of stated choice experiments: the state of practice and future challenges', working paper, ITS-WP-04-09, Institute of Transport Studies, University of Sydney.
Rose, J.M. and M.C.J. Bliemer (2005a), 'Constructing efficient choice experiments', working paper, ITS-WP-05-07, Institute of Transport Studies, University of Sydney.
Rose, J.M. and M.C.J. Bliemer (2005b), 'Sample optimality in the design of stated choice experiments', working paper, ITS-WP-05-13, Institute of Transport Studies, University of Sydney.
Rose, J.M. and M.C.J. Bliemer (2006), 'Designing efficient data for stated choice experiments', 11th International Conference on Travel Behaviour Research, Kyoto, 16–20 August.
Rose, J.M. and M.C.J. Bliemer (2009), 'Constructing efficient stated choice experimental designs', *Transport Reviews*, **29** (5), 587–617.
Rose, J.M., S. Bain and M.C.J. Bliemer (2011), 'Experimental design strategies for stated preference studies

dealing with non market goods', in J. Bennett (ed.), *International Handbook on Non-Marketed Environmental Valuation*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 273–9.
Rose, J.M., M.C.J. Bliemer, D.A. Hensher and A.C. Collins (2008), 'Designing efficient stated choice experiments involving respondent based reference alternatives', *Transportation Research Part B*, **42** (4), 395–406.
Rose, J.M., R. Scarpa and M.C.J. Bliemer (2009), 'Incorporating model uncertainty into the generation of efficient stated choice experiments: a model averaging approach', International Choice Modelling Conference, Harrowgate, 30 March–1 April.
Rousseas, S.W. and A.G. Hart (1951), 'Experimental verification of a composite indifference map', *Journal of Political Economy*, **59** (4), 288–318.
Sándor, Z. and M. Wedel (2001), 'Designing conjoint choice experiments using managers' prior beliefs', *Journal of Marketing Research*, **38** (4), 430–44.
Sándor, Z. and M. Wedel (2002), 'Profile construction in experimental choice designs for mixed logit models', *Marketing Science*, **21** (4), 455–75.
Sándor, Z. and M. Wedel (2005), 'Heterogeneous conjoint choice designs', *Journal of Marketing Research*, **42** (2), 210–18.
Scarpa, R. and J.M. Rose (2008), 'Designs efficiency for non-market valuation with choice modelling: how to measure it, what to report and why', *Australian Journal of Agricultural and Resource Economics*, **52** (3), 253–82.
Street, D.J. and L. Burgess (2004), 'Optimal and near optimal pairs for the estimation of effects in 2-level choice experiments', *Journal of Statistical Planning and Inference*, **118** (1–2), 185–99.
Street, D.J. and L. Burgess (2007), *The Construction of Optimal Stated Choice Experiments: Theory and Methods*, Hoboken, NJ: Wiley.
Street, D.J., D.S. Bunch and B.J. Moore (2001), 'Optimal designs for 2^k paired comparison experiments', *Communications in Statistics, Theory, and Methods*, **30** (10), 2149–71.
Street, D.J., L. Burgess and J.J. Louviere (2005), 'Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments', *International Journal of Research in Marketing*, **22** (4), 459–70.
Thurstone, L. (1931), 'The indifference function', *Journal of Social Psychology*, **2** (2), 139–67.
Toner, J.P., S.D. Clark, S.M. Grant-Muller and A.S. Fowkes (1998), 'Anything you can do, we can do better: a provocative introduction to a new approach to stated preference design', *WCTR Proceedings*, Antwerp, **3**, 107–20.
Toner, J.P., M. Wardman and G. Whelan (1999), 'Testing recent advances in stated preference design', *Proceedings of the European Transport Conference*, Cambridge.
Train, K. (2009), *Discrete Choice Methods with Simulation*, 2nd edn, Cambridge: Cambridge University Press.
Vermeulen, B., P. Goos and M. Vandebroek (2008), 'Models and optimal designs for conjoint choice experiments including a no-choice option', *International Journal of Research in Marketing*, **25** (2), 94–103.
Vermeulen, B., P. Goos, R. Scarpa and M. Vandebroek (2011a), 'Bayesian conjoint choice designs for measuring willingness to pay', *Environmental and Resource Economics*, **48** (1), 129–49.
Vermeulen, B., P. Goos, R. Scarpa and M. Vandebroek (2011b), 'Rank-order conjoint experiments: efficiency and design', *Journal of Statistical Planning and Inference*, **141** (8), 2519–31.
Watson, S.M., J.P. Toner, A.S. Fowkes and M.R. Wardman (2000), 'Efficiency properties of orthogonal stated preference designs', in J. de D. Ortuzar (ed.), *Stated Preference Modelling Techniques*, London: PTRC Education and Research Services Limited, pp. 91–101.
Yu, J., P. Goos and M. Vandebroek (2008), 'Model-robust design of conjoint choice experiments', *Communications in Statistics: Simulation and Computation*, **37** (8), 1603–21.
Yu, J., P. Goos and M. Vandebroek (2009), 'Efficient conjoint choice designs in the presence of respondent heterogeneity', *Marketing Science*, **28** (1), 122–35.
Yu, J., P. Goos and M. Vandebroek (2010), 'Comparing different sampling schemes for approximating the integrals involved in the efficient design of stated choice experiments', *Transportation Research Part B*, **44** (10), 1268–89.
Yu, J., P. Goos and M. Vandebroek (2011), 'Individually adapted sequential Bayesian designs for conjoint choice experiments', *International Journal of Research in Marketing*, **28** (4), 378–88.
Yu, J., P. Goos, P. and M. Vandebroek (2012), 'A comparison of different Bayesian design criteria to compute efficient conjoint choice experiments', *Transportation Research Part B*, **46** (7), 789–807.