**7**

**Sample optimality in the design of stated choice experiments**

***John M. Rose***

*University of Sydney, Sydney, AUSTRALIA*

***Michiel C.J. Bliemer***

*University of Sydney, Sydney, AUSTRALIA*

**Abstract**

Recent research by Bliemer and Rose (2005; 2009; 2010) and Rose and Bliemer (2005) suggest as a measure for calculating sample size requirements for models estimated using stated choice data, the *S*-error statistic. Prior to this, existing sampling theories failed to adequately address the issue of sample size requirements specifically for this type of data and hence researchers have had to resort to simple rules of thumb or ignore the issue and collect samples of arbitrary size, hoping that the sample is sufficiently large enough to produce reliable parameter estimates. In this paper, we explore the sample size calculations proposed by Bliemer and Rose demonstrate how these measures may be used to suggest a theoretical minimum sample size, assuming prior parameter values used in generating experiments. Sample size requirements for different model types are explored via three different case studies. The paper finds that the *S*-error statistic provides a robust estimate of the minimum sample size requirements for stated choice studies, however it is recommended that larger sample sizes than suggested by the statistic be collected to allow for different sources of misspecification that can occur during the course of such studies.

1. **Introduction**

The purpose behind conducting experiments is to determine the independent influence that different variables (called attributes or factors, depending on the literature cited) have upon some observed outcome. In the context of stated choice experiments (SCE), this means attempting to determine what influence the attributes of the SCE have upon the observed choices captured during the course of the experiment. Rather than observe the choices of a single, or even a limited number of respondents, it is typically necessary to pool the responses obtained from multiple respondents in order to produce statistically reliable parameter estimates. As such, studies utilising SCEs typically result in data sets made up of responses from a large number of respondents who are asked to complete more than one tasks involving the choice of alternatives from amongst a finite set of listed options. In each task, the alternatives, whether labelled or unlabelled, are typically defined on a number of different attribute dimensions, each of which are further described by pre-specified levels drawn from some underlying *experimental design*. The number of choice tasks each respondent is asked to undertake will generally be up to the total number of choice tasks drawn from the experimental design. Consequently, an archetypal SCE might require choice data be collected on 200 respondents, each of whom are observed to make eight choices, thus producing a total of 1600 choice observations.

The necessity to pool data has lead several authors to seek ways to reduce the number of choice observations necessary for reliable analysis of choice data (e.g., Huber and Zwerina, 1996; Sándor and Wedel, 2001; 2002; 2005; Kanninen, 2002; Bliemer and Rose, 2006). Primarily, these research efforts have attempted to produce more statistically efficient experimental designs which for a given level of accuracy, allow for either a reduction in the number of choice tasks shown to individual respondents or alternatively, a reduction in the number of respondents required to complete the experiment. Such designs have been widely studied within the literature. For example, Bunch, *et al*, (1996) studied statistically efficient main effects designs whilst Anderson and Wiley (1992) and Lazari and Anderson (1994) introduce methods to generate statistically efficient cross-effect designs.

Recently, Bliemer and Rose (2005; 2009; 2010) and Rose and Bliemer (2005) demonstrated how it is possible to analytically calculate a theoretical minimum sample size for any given SCE. Using the expected asymptotic variance covariance (AVC) matrix generated for a SCE, Bliemer and Rose showed that there exists a relationship between the expected standard errors of a design and the sample size requirements for that design. Bliemer and Rose further demonstrated how this relationship can be manipulated to provide an indication as to what sample size will be required for each parameter estimate related to the attributes of the design to be statistically significant. In doing so, Bliemer and Rose were able to derive a statistical measure, which they term *S*-error, which can be used to generate a design that will minimise the theoretical minimum sample size required of that SCE.

Whilst the *S*-error provides an analytical indication of the minimum sample size required for SCE, Bliemer and Rose (2005; 2009; 2010) found that depending on the assumptions made in generating the design, sample sizes of 10 or less may theoretically provide statically significant parameter estimates in some cases. Such sample sizes are unlikely to work in practice however given that estimation procedures often require substantially more observations in order to yield stable parameter estimates. Thus, whilst a design may theoretically yield low standard error values at a given small finite sample size, the question remains as to whether the parameter estimates themselves are stable in such small samples. If this is not the case, then reliance on *S*-error as an indicator of minimum sample size may result in biased study results.

To add an additional complication in terms of calculating the sample size requirements for SCEs, the unique nature of data collected from SCEs has implications in terms of the econometric outputs obtained during model estimation, implications that directly affect the sample size calculation itself. Most econometric models used in discrete choice studies fail to account for the nature of the repeated observations of SCE data, treating each observation as if it were made by a different individual (Ortúzar and Willumsen, 2001). Traditionally, the fact that multiple observations are typically captured from the same individual was thought to impact only upon the asymptotic standard errors derived from the variance covariance (VC) matrix of the model. Early research demonstrated that this problem could subsequently be corrected post estimation using either bootstrap or jackknife methods (e.g., Cirillo, *et al*, 2000 and Ortúzar, *et al*, 2000). Whilst bootstrap and jackknife techniques are still used in many empirical contexts, particularly when closed form econometric models are used, more recently, analysts have employed a variation of the log-likelihood specification of the error component or mixed multinomial logit (MMNL) model. Using this particular log-likelihood specification, the unobserved or random taste heterogeneity resident in SCE data is treated as varying across respondents but not over choice observations within respondents (see Revelt and Train, 1998; Train, 2003). By specifying the log-likelihood function in this way, the model directly accounts for the repeated observation issue of data collected from SCEs, not just in terms of the standard errors, but also in terms of potential biases in the parameter estimates (see Hess and Rose, 2009).

The issue for generating experimental designs, and in particular for calculating the sample size requirements for such designs, therefore becomes one of what model specification to assume in calculating the AVC matrix for the design. In generating the design, the idea is to calculate the AVC matrix so that it will look as close to possible to the VC matrix of the most likely model to be estimated on data collected using that design. The problem then becomes, if one is going to adjust the VC matrix post model estimation, via bootstrap or jackknife techniques, then the AVC matrix of the design should be similarly adjusted. This in turn will have an impact upon the sample size calculations for the design. To overcome this problem, we utilise the MMNL model form in this paper, however we do examine the impact of sample size on MNL models without correcting for potential biases resulting from biased standard errors. In doing so, we note that such corrections remain in the minority of reported studies anyway, and hence our analysis remains consistent with the majority of the published papers using the MNL model form.

The purpose of this paper is to explore the issue of sample size requirements for SCE further. In it, we explore the issue of sample size, using both Monte Carlo simulations as well as real life data to test the veracity of the *S*-efficiency measure to determine the sample sizes required for SCE. Using both the simulated and real life empirical data, and bootstrapping at different sample sizes, we are able to i) test the performance of the *S*-efficiency criteria at varying sample sizes and ii) examine at what sample size the parameter estimates obtained from different types of discrete choice models (i.e., MNL, MMNL and error components (EC) models) become stable. In doing so, we provide guidelines to transport researchers using SCE as to how to calculate the minimum sample sizes that may be required, as well as provide researchers with the tools to calculate these requirements for specific studies.

The remainder of the paper is organised as follows. In the following section, we describe the current methods used to construct SCEs. In doing so, we re-present the S-efficiency measure proposed by Bliemer and Rose (2005; 2009; 2010) and Rose and Bliemer (2005). In Section 3, we discuss the use of Monte Carlo simulation and bootstrapping studies which are used on three case studies, two using simulated data, and a third using real life empirical data, which are outlined in Section 4. Section 5 presents the results of the three case studies, before Section 6 provides discussion and conclusions.

1. **Stated Choice Experiments**

There exist numerous approaches to generate workable experimental designs. As such, the specific approach adopted by the researcher will reflect their beliefs as to what are the most important properties that SCE display. Before we are able to discuss the issue of calculating sample size requirements for SCE, it is first necessary to discuss the main approaches used today. We do this now.

**2.1 Traditional Orthogonal Designs Methods**

The most common experimental design type has been the *orthogonal design*. Orthogonality refers to the correlation structure of the design attributes where designs with zero correlations are said to be *orthogonal*. Under this definition, each attribute of the design will be independent of all other attributes, although some orthogonal designs assume orthogonality only for attributes within alternatives, but not between. Several methods for constructing orthogonal designs exist, such as balanced incomplete blocked designs (BIBD), Latin Squares designs, orthogonal in the differences fractional factorial designs, and fold-over designs. Such designs have been discussed elsewhere and hence are not discussed further here (see e.g., Bunch, *et al*, 1996; Fowkes and Wardman, 1988; Louviere, *et al*, 2000 or Rose and Bliemer, 2008).

The most common type of orthogonal design that is used in practice is known as the *LKJ* orthogonal fractional factorial design (where *L* is the number of levels, *K* the number of attributes and *J* the number of alternatives). Two types of *LKJ* orthogonal fractional factorial designs persist within the literature. The first type, known as a simultaneous orthogonal design, involves generating a design that is orthogonal both within and between alternatives. The second type of *LKJ* orthogonal fractional factorial design, known as a sequential orthogonal design, involves first generating an orthogonal design for the first alternative, and then using the same design to construct subsequent alternatives by re-arranging the rows of the design (see e.g., Louviere, *et al*, 2000). Independent of the actual process used, a number of useful websites and software are available for obtaining orthogonal designs. Sloane provides many orthogonal arrays online associated with Hedayat, *et al*, (1999). Several software packages such as SPSS (www.spss.com), SAS (www.sas.com*)* and Ngene (www.choice-metrics.com) are also able to generate a range of orthogonal designs.

**2.2 D-Optimal Design Method Under the Null Hypothesis**

A number of researchers have recently examined how to construct optimal sequential orthogonal designs under the assumption that the parameter estimates are zero. These designs maintain (within alternative) orthogonality, whilst also minimizing the elements of the AVC matrix assuming that the parameters will be zero and that the attributes will be orthonormally coded. In practice, this typically results in designs where the attribute levels across alternatives are made to be as different as possible. This means that this class of design will generally increase the trade-offs that respondents are forced to make across all attributes maximising the information obtained in terms of the importance that each attribute plays on choice (see e.g., Burgess and Street, 2005; Street and Burgess, 2004; 2007; Street, *et al*, 2005). Street and Burgess (2007), Street, *et al*, (2005) and Rose and Bliemer (2008) provide detailed discussions of the exact procedures used in generating this class of design, however Burgess (2007) and Ngene provide computing capabilities for automatically generating such designs[[1]](#footnote-1). For the purpose of this paper, we do not use this class of designs as they do not provide any information as to the likely sample sizes required.

**2.3 D-Efficient Design Methods Under the Non-Null Hypothesis**

An alternative approach to generating SCE involves selecting a design that is likely to provide an AVC matrix containing values which are as small as possible under the assumption that the parameters will be non-zero. Given that the asymptotic standard errors of the parameters obtained from discrete choice models are the square roots of the leading diagonal of the AVC matrix, the smaller the elements of the AVC matrix, the smaller the asymptotic standard errors for each of the parameters will be. Given that dividing the parameter estimates by the asymptotic standard errors produces the asymptotic *t*-ratios for the model, the smaller the asymptotic standard errors, the larger the asymptotic *t*-ratios for each of the parameter estimates obtained from the model will be. Designs which minimise the elements of the AVC matrix are referred to as *efficient* designs, which we note are unlikely to be orthogonal.

Efficient designs constructed under the non-null parameter prior hypothesis require a number of assumptions in order to construct the design. Firstly, given that the AVC matrix of one discrete choice model will differ to that of any other discrete choice model. For example, the AVC matrix of the MNL model is different to that of a nested logit or mixed multinomial logit (MMNL) model (see e.g., Bliemer, *et al*, 2009a; Bliemer and Rose, 2010; Rose, *et al*, 2009), it is necessary to first decide what model type is likely to be estimated once data has been collected. In this way, the AVC matrix can be generated for a design only under the assumption of a specific model type (although Rose, *et al*, 2009 use a model averaging approach to account simultaneously for multiple possible model types). Secondly, as the AVC matrix of any logit model is analytically equal to the negative inverse of the model’s Hessian of the log-likelihood function. This in turn is a function of the model probabilities. Given that the model probabilities are a function of the utilities which are in turn are a function of the design attributes and the parameter estimates, in order to predict the AVC matrix for a design, the researcher is required to assume what the population parameter estimates will be.

Fortunately, the researcher may assume prior parameter estimates in a Bayesian-like fashion when constructing the design. Precise prior parameter values need not be provided (though such designs have been generated in the past; see for example, Carlsson and Martinsson, 2002). Rather, prior parameter distributions that (hopefully) contain the true population parameter values might be considered. Such designs are then optimised over a range of possible parameter values, without the analyst having to know the precise population value in advance (see e.g., Sándor and Wedel, 2001 and Kessels, *et al*, 2006). This however increases substantially the computing time required to generate the design, see e.g., Bliemer, *et al*, (2008). Rose and Bliemer (2008) outline the precise steps used to generate this type of design whilst Bliemer and Rose (2010), Bliemer, *et al*, (2009a) and Rose, *et al*, (2009) provide details of the analytical second derivatives for a range of different logit models.

2.3.1 Measuring Statistical Efficiency. A number of measures of the statistical efficiency of a design have been proposed within the literature. The most commonly used measure is the *D*-*error* measure which uses the scaled determinant of the AVC to measure efficiency (scaled to account for the number of parameters to be estimated). The determinant of a matrix is a single summary statistic of the magnitude of the elements contained within the matrix. The smaller the determinant, the smaller, on average, the values contained within the matrix will be. In the case of designs generated under the null hypothesis assuming orthonormal coding and an MNL model structure, the *D*-*error measure* is converted to a *D*-optimality statistic which is a percentage value of the designs overall efficiency. For all other efficient designs, the objective is to minimise the *D*-*error* of the design. Similar to the *D*-*error*, some researchers prefer to use what is known as the *A*-*error*, which is calculated off of the trace of the AVC matrix (also normalised to account for the number of parameter estimates).

Another measure of statistical efficiency proposed by Bliemer and Rose (2005; 2009; 2010) is the *S*-error. This measure relies on the direct relationship that exists between the AVC matrix of logit models and sample size required to locate statistically significant parameter estimates, which was first described by McFadden (1974). McFadden (1974) demonstrated that the AVC matrices of discrete choice models are divisible by *N*, the sample size, and as such, the asymptotic standard errors are also divisible by the square root of *N*. Bliemer and Rose (2005; 2009; 2010) and Rose and Bliemer (2005) proposed exploiting this relationship to calculate the sample size requirements for SCE. The *S*-*error* of a design provides the theoretical minimum sample size required in order to obtain asymptotically significant parameter estimates for a design. As with the *D*-*error*, the objective is to find a design that minimises the *S*-*error* value. In order to calculate the *S*-error of a design, the analyst must also construct the AVC matrix for the design.

Independent of the precise efficiency measure used, minimizing the elements of the AVC matrix for a design also reduces the expected asymptotic standard errors (i.e., the square roots of the diagonals of the AVC matrix). As such, for any given sample size, smaller asymptotic standard errors mean smaller confidence intervals around the parameters estimates as well as larger asymptotic *t*-ratios for each of the parameters. Hence, efficient designs are constructed specifically for the purpose of producing more reliable parameter results. Alternatively, efficient designs may produce the same asymptotic standard errors as other designs given smaller sample sizes. Given that the AVC matrices of discrete choice models are divisible by *N* (assuming a fixed design), there exists an inescapable diminishing return in terms of the statistical significance of the parameter estimates obtained from each additional respondent added to a survey. Let the Fisher information matrix with *N* respondents be denoted by IN(β) and let ΩN(β) be the corresponding AVC matrix, given some (prior) parameter values β. Since IN(β)=N·I1(β), it holds that ΩN(β) = (IN(β))-1 = 1/N(I1(β))-1 = 1/NΩ1(β) such that with a sample size of *N* the (asymptotic) standard error for parameter *k* is

|  |  |
| --- | --- |
|  | (1) |

From Equation (1), it is clear that the asymptotic standard errors provide diminishing improvements (decreases) as the sample size increases. The importance of this relationship becomes clear when one considers the further relationship between the standard errors and the equation typically used to determine the statistical significance of the parameter estimates, the asymptotic *t*-ratio. Given that the asymptotic *t*-ratio for parameter *k* is given as

|  |  |
| --- | --- |
|  | (2) |

It is possible to re-arrange to give

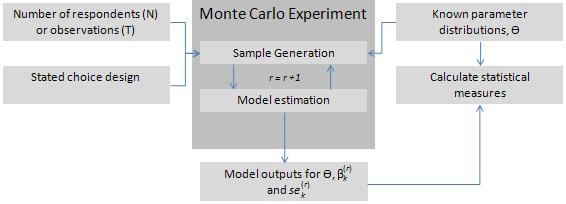
|  |  |
| --- | --- |
|  | (3) |

Under a non-null set of prior parameters, it is therefore possible to calculate the sample size *N* that will return a given *t* value. It is this relationship which Bliemer and Rose use in their *S*-error calculation. In this paper, for demonstration purposes, we will mostly use a *t* value of 1.96. Note that the analyst can set this value, where higher values yield more reliable parameter estimates (and also larger sample sizes).

1. **Monte Carlo Simulations**

In order to test the sample size requirements against predicted sample sizes, we construct a number of different case studies. As part of this paper, we present three different case studies, two utilising simulated data and the third real empirical data. In order to reduce bias in the data generation process for the simulated case studies, we start by generating the simulated data and then, with the attributes already fixed, simulate different respondents performing the choice tasks. This process ensures that the attributes and attribute levels are held constant within each simulated data set. Monte Carlo simulations are then performed on each of the data sets to compare the observed model results against the known underlying behavioural rules.

Figure 1 demonstrates the Monte Carlo simulation process employed for the current study. Starting with the generated experimental design, data based on a sample of respondents (or choice observations) is first simulated, including a choice variable, ynsj, which is an indicator equal to one if respondent *n* chooses alternative *j* in choice situation *s*, and zero otherwise. The choice variable is created by first calculating the utilities for each choice observation based on the design attributes and a set of random parameter draws[[2]](#footnote-2) from given parameter distributions (where the specific set of draws is kept fixed across choice tasks for the same respondent), which when added to an additional simulated random error term (EV1 distributed), produces different utility values for each of the alternatives. Once the utilities for each alternative are calculated, the alternative with the highest utility is assumed to be the one selected. By simulating a large number of respondents, we are able to draw different subsets of respondents, a total of *R* times, from which MNL, MMNL and Error components models are estimated with the resulting outputs (e.g., parameter values, standard errors). These model outputs are then compared to the known input values (e.g., parameter priors, expected standard errors).



*Figure 1. Monte Carlo simulation process*

In order to test sample size influences, we incrementally increase the number of respondents drawn from the generated sample. We then examine the resulting model outputs, including the parameter estimates and parameter *t*-ratios over the different Monte Carlo simulations. A number of different measures have been proposed to compare the known model inputs to those obtained from the Monte Carlo simulation process. The two most popular statistical methods are the mean square error (MSE) and the relative absolute error (RAE). Equations for the calculating the MSE and RAE statistics are given in Equations (4) and (5) respectively for the parameter estimates. It is possible to produce similar measures comparing the asymptotic *t*-ratios given the expected asymptotic *t*-ratios calculated using Equation (2).

|  |  |
| --- | --- |
|  | (4) |
|  | (5) |

Where is the parameter estimate obtained at sample iteration *r*, and is the known prior parameter estimate used in constructing the Monte Carlo simulation.

In addition to using simulated data, we also utilise a real empirical data set to test sample size requirements for SCEs. In using this data set, a similar process to that described above is used to test the model outputs under various sample sizes. Rather than use the simulated choice variable, ynsj, however, the real observed choices are used instead. Subsamples of the data are then randomly selected and similar calculations are estimated to test the model outputs against expected outcomes.

1. **Case Studies**

In this section, we discuss three case studies, two of which utilize data simulated for different SCEs. In addition to the simulated data, we present a third case study that utilizes real data. In the first simulated case study, we compare and contrast different experimental designs; the first represents an *LKJ* orthogonal design and the second a design optimized for a MMNL model with generic parameter estimates. In the second simulated case, we examine the impact upon model outputs at different sample sizes for a single fixed design but allowing for utility specifications than assumed in the first case study. The final study, utilizing an empirical data set, like the second case study, compares sample size measures for a given design type. In the first two case studies, sample size requirements are examined under different model specification, whilst the last case study explores sample size issues for the MMNL model only.

**4.1 Case Study 1: Simulated Data 1**

Consider a SCE involving three alternatives, the first two of which are described by four attributes, and the last representing a no choice or status quo alternative and hence having no associated attributes. For simplicity we assume that all attribute related parameters are generic, although the application also assumes alternative specific constants. In generating the design, we assume that each respondent will engage in 16 choice tasks. Within the SCE, the attribute levels can take on different values for each of the 16 choice tasks shown to respondents. Let us assume that the first three attributes can take on one of four levels with these being {5, 10,15, 20} for the first attribute and {0, 1, 2, 3} for the second and third attributes, whilst the last attribute can take two levels, {0, 1} The generated designs assume attribute level balance (see Bliemer and Rose, 2006).

Equation (6) shows the utility specification set up used for the case study.

|  |  |
| --- | --- |
|  | (6) |

Where ηA,B ~ N(0,σ2) represents an EC model specification.

Two designs are used as part of the case study. The first is an *LKJ* orthogonal fractional factorial design whilst the second is a *D*-efficient design generated under the non-null hypothesis assuming a MMNL model specification. The latter design requires that priors be assumed in generating the design. Table 1 shows the priors assumed for the parameter and error component estimates used not only in generating the *D*-efficient design, but also used for the Monte Carlo simulation process.

*Table 1: Case study and Monte Carlo simulation parameter priors*

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Moment** | **Prior value** |
| β0A | Mean | -2.4 |
| β0B | Mean | -2.2 |
| β1 | Mean | N(-0.08,0.01) |
| Std Dev. | U(0.02,0.04) |
| β2 | Mean | 1.2 |
| Std Dev. | 0.4 |
| β3 | Mean | 1.2 |
| β4 | Mean | -0.7 |

In specifying the priors for generating the *D*-efficient design, we have assumed the first two parameters to be random parameters drawn from a Multivariate Normal distribution, thus representing unimodal and symmetric preference heterogeneity over the sampled population for these two attributes. In specifying the prior parameter values for these distributions, we have assumed a Bayesian design generation approach for the first random parameter, with the mean of the random parameter distribution being drawn from a Normal distribution and the standard deviation prior from a uniform distribution. In this way, we have incorporated uncertainty as to the precise values both population moments will take post data collection for this random parameter. For the second random parameter distribution, we have assumed fixed or known mean and standard deviation parameter priors for the design generation process. The two designs are given in Table 2.

To test the sample size requirements for the different designs, we assume three different model specifications during the Monte Carlo simulation process. In addition to assuming

*Table 2. Case study 1 designs*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Orthogonal** | | | | | | | |
| **S** | ***x*1A** | ***x*2A** | ***x*3A** | ***x*4A** | ***x*1B** | ***x*2B** | ***x*3B** | ***x*4B** |
| 1 | 5 | 0 | 0 | 1 | 5 | 0 | 0 | 1 |
| 2 | 15 | 1 | 3 | 1 | 20 | 0 | 2 | 1 |
| 3 | 10 | 1 | 2 | 0 | 5 | 1 | 2 | 0 |
| 4 | 20 | 0 | 1 | 0 | 20 | 1 | 0 | 0 |
| 5 | 10 | 3 | 1 | 1 | 15 | 0 | 3 | 0 |
| 6 | 20 | 2 | 2 | 1 | 10 | 0 | 1 | 0 |
| 7 | 5 | 2 | 3 | 0 | 15 | 1 | 1 | 1 |
| 8 | 15 | 3 | 0 | 0 | 10 | 1 | 3 | 1 |
| 9 | 15 | 2 | 2 | 0 | 15 | 2 | 2 | 0 |
| 10 | 20 | 1 | 3 | 0 | 5 | 2 | 1 | 1 |
| 11 | 15 | 3 | 1 | 1 | 5 | 3 | 0 | 0 |
| 12 | 20 | 0 | 0 | 1 | 15 | 3 | 3 | 1 |
| 13 | 10 | 2 | 1 | 0 | 10 | 2 | 2 | 1 |
| 14 | 5 | 1 | 0 | 0 | 20 | 2 | 1 | 0 |
| 15 | 10 | 3 | 2 | 1 | 20 | 3 | 0 | 1 |
| 16 | 5 | 0 | 3 | 1 | 10 | 3 | 3 | 0 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **MMNL (panel)** | | | | | | | |
| **S** | ***x*1A** | ***x*2A** | ***x*3A** | ***x*4A** | ***x*1B** | ***x*2B** | ***x*3B** | ***x*4B** |
| 1 | 5 | 3 | 0 | 1 | 20 | 0 | 3 | 0 |
| 2 | 5 | 0 | 3 | 0 | 15 | 2 | 1 | 1 |
| 3 | 15 | 2 | 0 | 1 | 10 | 1 | 0 | 0 |
| 4 | 15 | 3 | 1 | 0 | 10 | 0 | 2 | 1 |
| 5 | 20 | 3 | 2 | 1 | 5 | 1 | 2 | 0 |
| 6 | 20 | 2 | 2 | 0 | 5 | 3 | 1 | 1 |
| 7 | 10 | 1 | 1 | 0 | 15 | 3 | 2 | 1 |
| 8 | 20 | 2 | 2 | 0 | 5 | 2 | 2 | 1 |
| 9 | 20 | 1 | 3 | 0 | 5 | 3 | 1 | 1 |
| 10 | 10 | 0 | 3 | 1 | 20 | 3 | 0 | 0 |
| 11 | 5 | 3 | 0 | 0 | 20 | 1 | 3 | 1 |
| 12 | 15 | 0 | 3 | 1 | 10 | 2 | 0 | 0 |
| 13 | 10 | 2 | 0 | 1 | 15 | 0 | 3 | 0 |
| 14 | 5 | 1 | 2 | 0 | 20 | 2 | 3 | 1 |
| 15 | 10 | 1 | 1 | 1 | 10 | 0 | 1 | 0 |
| 16 | 15 | 0 | 1 | 1 | 15 | 1 | 0 | 0 |

a MMNL model with the priors as given in Table 1, we also assume MNL and EC specifications. The choice vector for the sample was generated assuming the MMNL model parameters used in the design generation phase. As such, the MNL and EC model specifications represent both model and prior parameter misspecification. To test sample size requirements, we began with assuming 25 respondents, which is then incrementally increased by 25 respondents up to a total of 250 respondents (i.e., the first simulation involved 25 respondents, the second 50, the third 75, etc.). For each model type and sample size, *R* = 100 Monte Carlo iterations are performed on the simulated data sets.

**4.2 Case Study 2: Simulated Data 2**

The second case study was constructed to demonstrate the impact of sample size requirements not just for generic parameter estimates, but also upon alternative specific parameter estimates beyond ASCs. The utility specification used in generating the design for the second case example is given in Equation (7).

|  |  |
| --- | --- |
|  | (7) |

For the first alternative, we include a fixed constant term. For the first and second attributes, generic parameters are assumed whilst for the third attribute of each alternative, the parameters are assumed to be alternative-specific. Parameters β1 and β4 are assumed to be random parameters, following normal distributions, βk ~ N(µk,σk), k=1,…,4 with prior values specified in Table 3. The design was generated with 12 choice tasks. The design used for the case study is reported in Table 4.

*Table 3. Case study and Monte Carlo simulation parameter priors*

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Moment** | **Prior value** |
| β0A | Mean | -0.5 |
| β1 | Mean | -0.7 |
| Std Dev. | 0.3 |
| β2 | Mean | -1.5 |
| β3 | Mean | -0.8 |
| β4 | Mean | -0.9 |
| Std Dev. | 0.2 |

*Table 4. Case study 2 design*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S** | ***x*1A** | ***x*2A** | ***x*3A** | ***x*1B** | ***x*2B** | ***x*3B** |
| 1 | 15 | 1 | 1 | 5 | 2 | 2 |
| 2 | 5 | 1 | 0 | 15 | 0 | 0 |
| 3 | 15 | 0 | 3 | 10 | 0 | 1 |
| 4 | 10 | 1 | 3 | 10 | 1 | 2 |
| 5 | 5 | 0 | 1 | 15 | 1 | 0 |
| 6 | 5 | 1 | 2 | 15 | 2 | 1 |
| 7 | 15 | 0 | 0 | 5 | 2 | 0 |
| 8 | 10 | 2 | 2 | 15 | 0 | 1 |
| 9 | 15 | 0 | 2 | 5 | 2 | 0 |
| 10 | 5 | 2 | 3 | 10 | 1 | 2 |
| 11 | 10 | 2 | 0 | 10 | 0 | 2 |
| 12 | 10 | 2 | 1 | 5 | 1 | 1 |

Similar to case study 1, we report the results from a number of Monte Carlo simulations on data generated using the above design and parameter estimates. Also similar to the first case study, we test several model specifications including the MNL and MMNL models. For the MMNL simulations, we use the same parameter estimates that were used to generate the design to construct the choice vector. Unlike the first case study, for the MNL model estimates, we use simulated choices generated for the MMNL model. As such, the MNL model results represent a test of model misspecification during the estimation phase. To test sample size requirements, we simulate for both model types, data beginning with five respondents and incrementally increase this by five additional respondents up to 120 respondents (i.e., we assume 5, 10, 15, etc. respondents. For each model type and sample size, *R* = 100 Monte Carlo iterations are performed.

**4.3 Case Study 3: Empirical Data Set**

The final case study involves the analysis of real data collected as part of a SCE conducted in the Netherlands in 2007. The case study involved respondents having to select their preferred alternative from amongst five hypothetical airline tickets given the scenario that they were travelling from Amsterdam to Barcelona for a holiday. Each alternative was described by six attributes, five of which were assigned three levels each with the last having six levels. The attributes and their respective levels are given in Table 5. Although not listed in the table, a seventh attribute was part of the design. This attribute representing airport destination acted as a switching variable, changing the attribute levels for egress price and travel time.

Three experimental designs were generated for the study, one orthogonal design with 108 choice tasks, one Bayesian D-efficient design with 108 choice tasks, and one Bayesian D-efficient design with 18 choice tasks. Each design was blocked so that individual respondents only completed six choice tasks each. For the current case study, we utilise only data collected from the last design with 18 choice tasks, which can be found in Appendix A. More information about the study may be found in Bliemer, *et al*, (2009b).

*Table 5. The attribute levels used as part of the DCE*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **(lvl) Airline** | **(lvl) Ticket Price** | **(lvl) Departure Time** | | **(lvl) Transfers** | **(lvl) Egress Price** | | **(lvl) Egress Time** | |
| (0) Air France | (0) €50 | (0) 6:00 | | (0) 1 Non stop | (0) €1\* or €9 | | (0) 20 min\* or 1 hr | |
| (1) KLM | (1) €75 | (1) 12:00 | | (1) 1 hour | (1) €3\* or €12 | | (1) 30 min\* or 1hr20 | |
| (2) Iberia | (2) €100 | (2) 18:00 | | (2) 2 hours | (2) €5\* or €15 | | (2) 40 min\* or 1hr40 | |
| (3) Vueling |  |  | |  |  | |  | |
| (4) Transavia |  |  | |  |  | |  | |
| (5) Easy Jet |  | |  | | |  | |  |

\* If flown via Barcelona Airport, other values are through Girona Airport

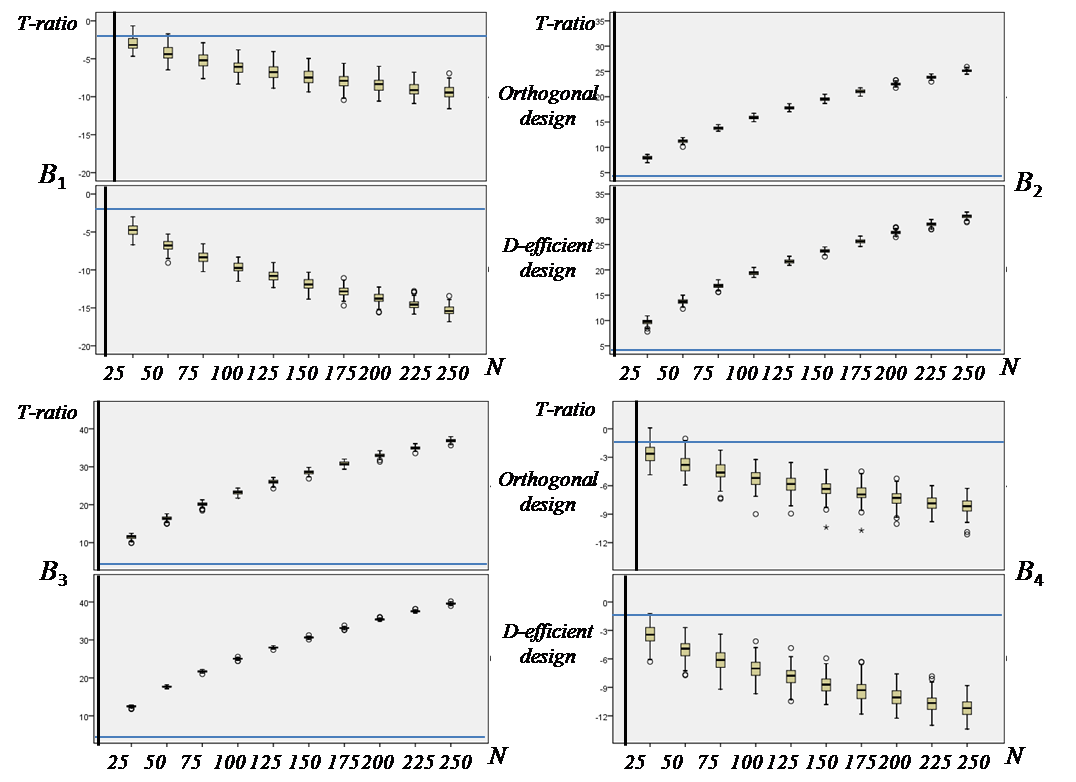
1. **Case Study Results**

In this section, we present the results of the simulation and bootstrapping exercises performed on each of the three case studies. In presenting the results, we look at the MSE and RAE results for the parameters and asymptotic *t*-ratios computed at different sample sizes to determine how well the designs perform. We then compare the design performance to predicted performance given *S*-error predictions.

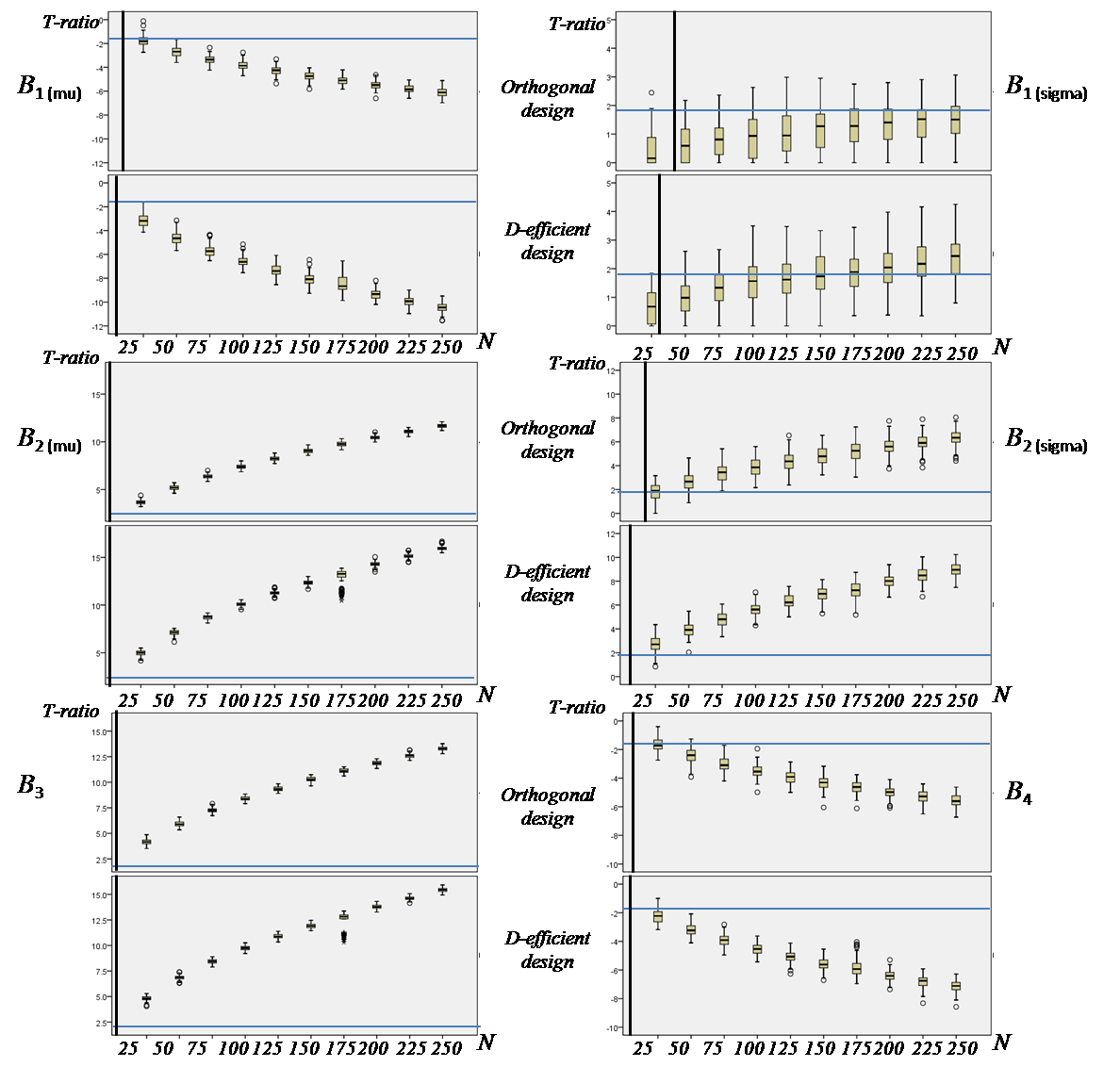
**5.1 Case Study 1: Simulated Data 1**

Figures 2–4 report boxplots of the *t*-ratios at different sample sizes obtained from the simulated data over the *R* =100 Monte Carlo simulations for the MMNL, MNL and EC models respectively. Horizontal lines on each graph represent a *t*-value of 1.96. Vertical lines shown represent the parameter specific *S*-error values determined during the design generation process. With the exception of the *S*-error of the standard deviation of the first random parameter, the *S*-error values do appear to represent a good approximation of the lower sample size bounds of the parameter estimates across both designs and all three model types. The sole exception to this, as stated above, is the *S*-error value for the standard deviation parameter associated with the first random parameter. Given that the *S*-error value for the second standard deviation parameter does appear to represent a good approximation of the necessary sample size required to achieve statistical significance, the likely reason for the failure of the calculation for this parameter lies in the magnitude of the parameter itself, which being only 0.03 is relatively close to zero. It is therefore possible that such small amounts of preference heterogeneity are difficult for the model to retrieve and hence larger sample sizes are necessary in order to derive enough statistical power to detect that the parameter itself is different from zero. Even if such a hypothesis is true, Bliemer and Rose (2010) state that the *S*-error calculation should only be used to determine a minimum theoretical sample size.

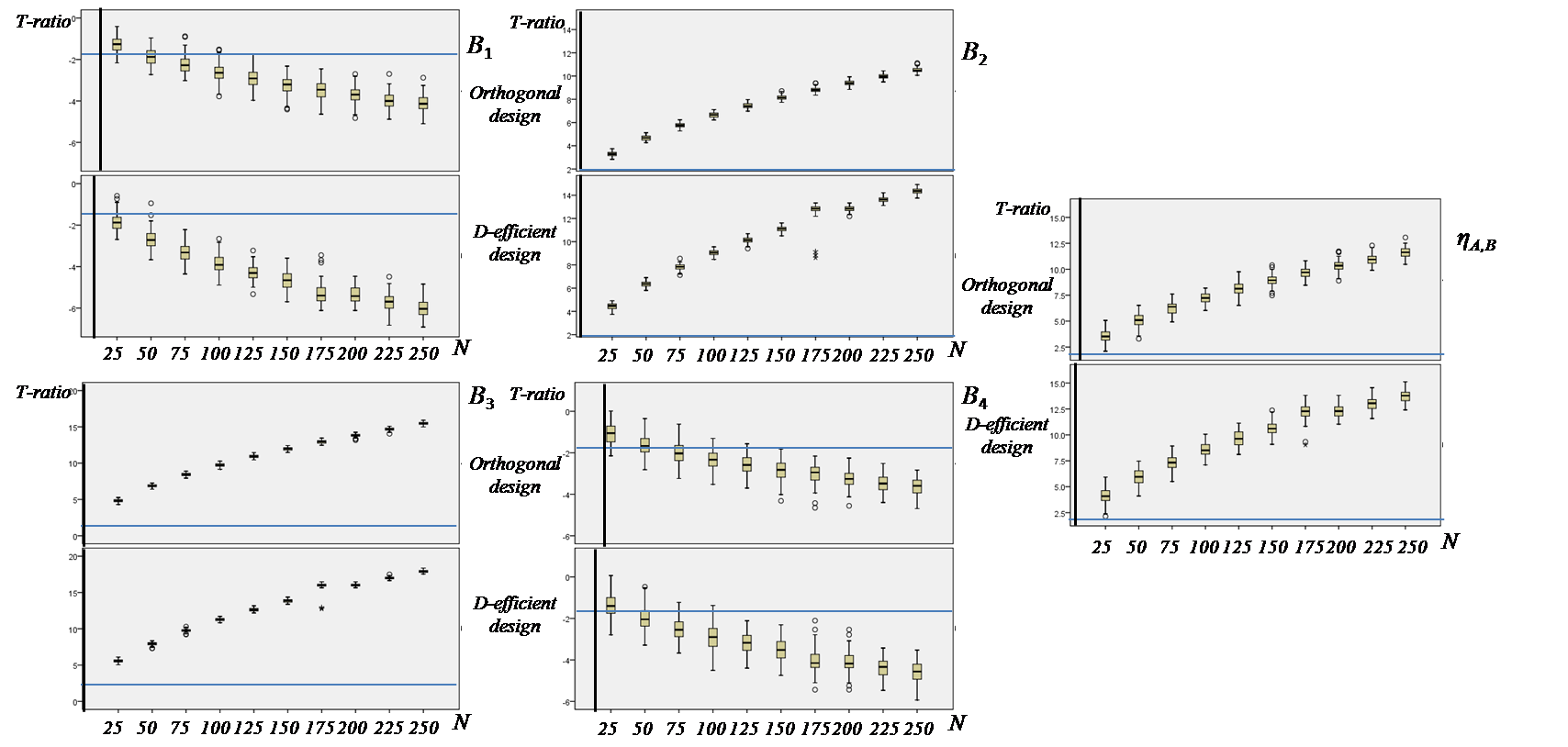
Tables 6–8 reports the MSE and RAE statistics for the two designs under the three different model specifications. Whilst as is to be expected, larger sample sizes produce less biased parameter estimates, what become clear from the table is that even after only 25 respondents, the parameter estimates tend to reproduce the known parameter values with minimal bias. Nevertheless, close examination of the table reveals substantial reductions in parameter biases moving from 25 to 50 respondents after which diminishing reductions in bias are observed. Although reductions in bias are observed after 100 respondents, such reductions appear to be marginal. Given this observation, whilst the *S*-error calculation does appear to provide evidence as to the theoretical minimum sample size required to obtain statistical significance, larger sample sizes may be required to reduce bias in the parameter estimates.



*Figure 2. Case study 1 MNL model t-ratio results*



*Figure 3. Case study 1 MMNL model t-ratio results*



*Figure 4. Case study 1 EC model t-ratio results*

*Table 6. Case Study 1 MSE and RAE MNL model results*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Orthogonal design** | | | | | **MMNL Db-efficient design** | | | | |
| **MSE (β)** | | | | | | | | | |
| N | β1 | β2 | β3 | β4 | N | β1 | β2 | β3 | β4 |
| 25 | 0.0006 | 0.029 | 0.0283 | 0.061 | 25 | 0.0002 | 0.019 | 0.0191 | 0.039 |
| 50 | 0.0003 | 0.012 | 0.0119 | 0.029 | 50 | 0.0001 | 0.008 | 0.0097 | 0.019 |
| 75 | 0.0002 | 0.008 | 0.0069 | 0.020 | 75 | 0.0001 | 0.005 | 0.0061 | 0.012 |
| 100 | 0.0002 | 0.007 | 0.0054 | 0.015 | 100 | 0.00004 | 0.003 | 0.0041 | 0.009 |
| 125 | 0.0001 | 0.006 | 0.0043 | 0.013 | 125 | 0.00003 | 0.002 | 0.0029 | 0.006 |
| 150 | 0.0001 | 0.005 | 0.0037 | 0.010 | 150 | 0.00003 | 0.002 | 0.0025 | 0.005 |
| 175 | 0.0001 | 0.004 | 0.0034 | 0.008 | 175 | 0.00002 | 0.001 | 0.0021 | 0.005 |
| 200 | 0.0001 | 0.003 | 0.0022 | 0.006 | 200 | 0.00002 | 0.001 | 0.0020 | 0.004 |
| 225 | 0.0001 | 0.002 | 0.0019 | 0.004 | 225 | 0.00002 | 0.001 | 0.0016 | 0.003 |
| 250 | 0.0001 | 0.002 | 0.0015 | 0.004 | 250 | 0.00002 | 0.001 | 0.0016 | 0.003 |
| **RAE (β)** | | | | | | | | | |
| 25 | 0.2859 | 0.111 | 0.0732 | 0.330 | 25 | 0.1544 | 0.089 | 0.0599 | 0.269 |
| 50 | 0.2082 | 0.071 | 0.0463 | 0.223 | 50 | 0.1051 | 0.059 | 0.0436 | 0.178 |
| 75 | 0.1686 | 0.058 | 0.0363 | 0.186 | 75 | 0.0856 | 0.046 | 0.0357 | 0.145 |
| 100 | 0.1441 | 0.056 | 0.0319 | 0.153 | 100 | 0.0717 | 0.035 | 0.0284 | 0.125 |
| 125 | 0.1336 | 0.050 | 0.0293 | 0.142 | 125 | 0.0629 | 0.028 | 0.0241 | 0.103 |
| 150 | 0.1225 | 0.045 | 0.0261 | 0.123 | 150 | 0.0652 | 0.028 | 0.0213 | 0.097 |
| 175 | 0.1113 | 0.043 | 0.0259 | 0.110 | 175 | 0.0560 | 0.025 | 0.0210 | 0.096 |
| 200 | 0.0955 | 0.035 | 0.0201 | 0.096 | 200 | 0.0536 | 0.021 | 0.0208 | 0.084 |
| 225 | 0.0886 | 0.030 | 0.0185 | 0.090 | 225 | 0.0475 | 0.022 | 0.0181 | 0.074 |
| 250 | 0.0853 | 0.029 | 0.0167 | 0.086 | 250 | 0.0497 | 0.019 | 0.0173 | 0.073 |

*Table 7. Case Study 1 MSE and RAE MMNL model results*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Orthogonal design** | | | | | | |
| **MSE (β)** | | | | | | |
| N | β1(mu) | β1(sigma) | β2mu) | β2(sigma) | β3 | β4 |
| 25 | 0.0004 | 0.081 | 0.022 | 0.017 | 0.013 | 0.048 |
| 50 | 0.0002 | 0.079 | 0.010 | 0.007 | 0.006 | 0.027 |
| 75 | 0.0001 | 0.079 | 0.006 | 0.005 | 0.003 | 0.016 |
| 100 | 0.0001 | 0.079 | 0.004 | 0.004 | 0.003 | 0.011 |
| 125 | 0.0001 | 0.078 | 0.004 | 0.004 | 0.002 | 0.009 |
| 150 | 0.0001 | 0.078 | 0.003 | 0.003 | 0.002 | 0.006 |
| 175 | 0.00005 | 0.078 | 0.003 | 0.003 | 0.002 | 0.005 |
| 200 | 0.00003 | 0.078 | 0.002 | 0.002 | 0.001 | 0.004 |
| 225 | 0.00003 | 0.078 | 0.002 | 0.002 | 0.001 | 0.004 |
| 250 | 0.00004 | 0.078 | 0.002 | 0.002 | 0.001 | 0.004 |
| **RAE (β)** | | | | | | |
| 25 | 0.200 | 0.944 | 0.100 | 0.255 | 0.078 | 0.245 |
| 50 | 0.142 | 0.937 | 0.065 | 0.170 | 0.053 | 0.189 |
| 75 | 0.090 | 0.937 | 0.048 | 0.143 | 0.038 | 0.141 |
| 100 | 0.086 | 0.934 | 0.044 | 0.136 | 0.035 | 0.116 |
| 125 | 0.084 | 0.933 | 0.042 | 0.123 | 0.029 | 0.107 |
| 150 | 0.073 | 0.931 | 0.038 | 0.107 | 0.030 | 0.091 |
| 175 | 0.069 | 0.930 | 0.036 | 0.101 | 0.029 | 0.085 |
| 200 | 0.060 | 0.930 | 0.032 | 0.093 | 0.024 | 0.072 |
| 225 | 0.057 | 0.929 | 0.032 | 0.090 | 0.025 | 0.074 |
| 250 | 0.055 | 0.928 | 0.028 | 0.078 | 0.022 | 0.072 |

*Table 7. Continued*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **MMNL Db-efficient design** | | | | | | |
| **MSE (β)** | | | | | | |
| N | β1(mu) | β1(sigma) | β2mu) | β2(sigma) | β3 | β4 |
| 25 | 0.00016 | 0.079 | 0.019 | 0.008 | 0.012 | 0.031 |
| 50 | 0.00009 | 0.078 | 0.009 | 0.004 | 0.006 | 0.013 |
| 75 | 0.00005 | 0.077 | 0.006 | 0.003 | 0.004 | 0.009 |
| 100 | 0.00004 | 0.077 | 0.003 | 0.002 | 0.003 | 0.006 |
| 125 | 0.00003 | 0.078 | 0.003 | 0.001 | 0.002 | 0.004 |
| 150 | 0.00003 | 0.077 | 0.002 | 0.001 | 0.002 | 0.003 |
| 175 | 0.00003 | 0.077 | 0.003 | 0.001 | 0.002 | 0.004 |
| 200 | 0.00002 | 0.077 | 0.002 | 0.001 | 0.001 | 0.002 |
| 225 | 0.00002 | 0.077 | 0.001 | 0.001 | 0.001 | 0.003 |
| 250 | 0.00002 | 0.077 | 0.001 | 0.001 | 0.001 | 0.002 |
| **RAE (β)** | | | | | | |
| 25 | 0.132 | 0.936 | 0.091 | 0.179 | 0.072 | 0.208 |
| 50 | 0.094 | 0.932 | 0.062 | 0.123 | 0.052 | 0.130 |
| 75 | 0.073 | 0.928 | 0.053 | 0.108 | 0.043 | 0.105 |
| 100 | 0.065 | 0.927 | 0.037 | 0.083 | 0.036 | 0.088 |
| 125 | 0.057 | 0.927 | 0.037 | 0.077 | 0.031 | 0.072 |
| 150 | 0.054 | 0.926 | 0.034 | 0.072 | 0.027 | 0.066 |
| 175 | 0.050 | 0.926 | 0.030 | 0.069 | 0.028 | 0.074 |
| 200 | 0.045 | 0.926 | 0.026 | 0.056 | 0.023 | 0.053 |
| 225 | 0.041 | 0.925 | 0.025 | 0.062 | 0.020 | 0.057 |
| 250 | 0.040 | 0.925 | 0.025 | 0.054 | 0.018 | 0.052 |

*Table 8. Case Study 1 MSE and RAE EC model results*

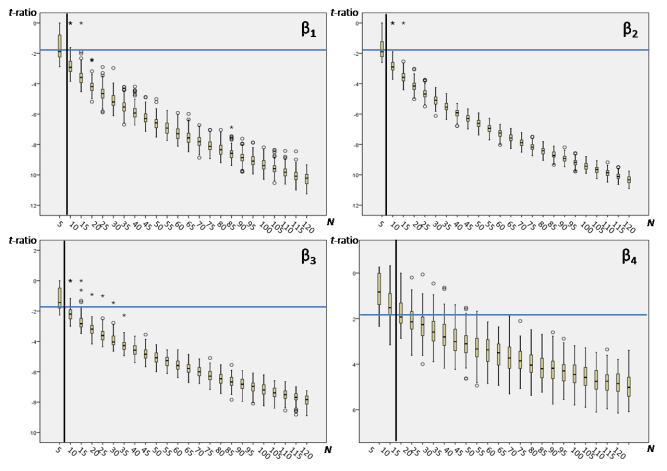
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Orthogonal design** | | | | | |
| **MSE (β)** | | | | | |
| N | β1 | β2 | β3 | β4 | ηA*B* |
| 25 | 0.0004 | 0.012 | 0.012 | 0.051 | 0.073 |
| 50 | 0.0002 | 0.006 | 0.006 | 0.024 | 0.044 |
| 75 | 0.0002 | 0.004 | 0.004 | 0.016 | 0.026 |
| 100 | 0.0001 | 0.003 | 0.003 | 0.010 | 0.020 |
| 125 | 0.0001 | 0.003 | 0.003 | 0.008 | 0.019 |
| 150 | 0.0001 | 0.002 | 0.002 | 0.006 | 0.016 |
| 175 | 0.0001 | 0.002 | 0.002 | 0.006 | 0.012 |
| 200 | 0.0001 | 0.002 | 0.002 | 0.004 | 0.012 |
| 225 | 0.00004 | 0.002 | 0.001 | 0.004 | 0.010 |
| 250 | 0.00005 | 0.001 | 0.001 | 0.004 | 0.009 |
| **RAE (β)** | | | | | |
| 25 | 0.273 | 0.088 | 0.058 | 0.363 | 0.150 |
| 50 | 0.184 | 0.063 | 0.040 | 0.244 | 0.109 |
| 75 | 0.163 | 0.050 | 0.035 | 0.204 | 0.086 |
| 100 | 0.133 | 0.048 | 0.031 | 0.165 | 0.076 |
| 125 | 0.127 | 0.043 | 0.027 | 0.145 | 0.077 |
| 150 | 0.119 | 0.040 | 0.025 | 0.131 | 0.064 |
| 175 | 0.115 | 0.039 | 0.025 | 0.123 | 0.059 |
| 200 | 0.100 | 0.032 | 0.021 | 0.102 | 0.060 |
| 225 | 0.084 | 0.030 | 0.021 | 0.101 | 0.057 |
| 250 | 0.093 | 0.029 | 0.020 | 0.099 | 0.052 |

*Table 8. Continued*

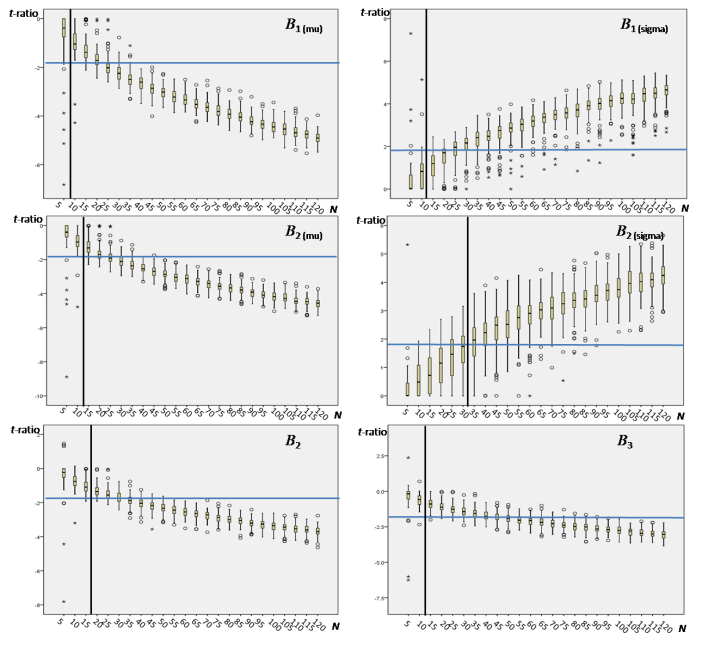
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **MMNL Db-efficient design** | | | | | |
| **MSE (β)** | | | | | |
| N | β1 | β2 | β3 | β4 | ηA*B* |
| 25 | 0.0002 | 0.011 | 0.017 | 0.046 | 0.091 |
| 50 | 0.0002 | 0.006 | 0.008 | 0.022 | 0.043 |
| 75 | 0.0001 | 0.004 | 0.005 | 0.013 | 0.026 |
| 100 | 0.0001 | 0.002 | 0.003 | 0.011 | 0.018 |
| 125 | 0.00005 | 0.002 | 0.002 | 0.007 | 0.018 |
| 150 | 0.00005 | 0.001 | 0.002 | 0.006 | 0.014 |
| 175 | 0.00003 | 0.001 | 0.002 | 0.004 | 0.010 |
| 200 | 0.00003 | 0.001 | 0.002 | 0.004 | 0.009 |
| 225 | 0.00003 | 0.001 | 0.001 | 0.003 | 0.009 |
| 250 | 0.00003 | 0.001 | 0.001 | 0.003 | 0.007 |
| **RAE (β)** | | | | | |
| 25 | 0.206 | 0.087 | 0.068 | 0.335 | 0.163 |
| 50 | 0.162 | 0.063 | 0.047 | 0.228 | 0.113 |
| 75 | 0.127 | 0.048 | 0.037 | 0.185 | 0.087 |
| 100 | 0.109 | 0.037 | 0.029 | 0.171 | 0.072 |
| 125 | 0.084 | 0.036 | 0.025 | 0.139 | 0.073 |
| 150 | 0.093 | 0.029 | 0.024 | 0.127 | 0.059 |
| 175 | 0.073 | 0.026 | 0.021 | 0.094 | 0.050 |
| 200 | 0.074 | 0.026 | 0.021 | 0.093 | 0.048 |
| 225 | 0.071 | 0.025 | 0.019 | 0.091 | 0.050 |
| 250 | 0.070 | 0.024 | 0.019 | 0.090 | 0.043 |

**5.2 Case Study 2: Simulated Data 2**

Figures 5 and 6 graph boxplots similar to those produced for the first case study. As with Figure 2–4, horizontal lines in Figures 5 and 6 represent *t*-ratios of 1.96 whilst vertical lines represent the calculated *S*-error statistics for each of the parameters shown. For reasons of space, results for *β*4 have been omitted. In omitting these results, we note that the conclusions drawn for this parameter do not differ from those of the other parameters and hence do not change the general findings of the case study. Examination of the performance of the *S*-error values for the parameters shown reveals that despite the model being misspecified, the predicted sample size requirements for the MNL model match very closely with the actual median required sample size. Nevertheless, the left hand side whiskers (i.e., the tails closer to zero) of the boxplots for the two true non random parameters (i.e., *β*3and*β*4) suggest that despite the median *t*-ratios for these parameters being close to that suggested by the predicted required sample sizes as given by the *S*-errors, there remains a considerable number of *t*-ratios that were not statistically significant at samples sizes even larger than the *S*-error. Hence, whilst the *S*-error performs well on average, the statistic remains a theoretical bound and should be interpreted as such. Examination of the MMNL model results produces similar findings. In this case, the plots reveal that the *S*-errors for the MMNL model tend to under calculate the sample size requirements by a factor of approximately two to four. Once more, the *S*-error calculation appears to produce an estimate of the theoretically minimum sample size that should be used in practice with larger sample sizes recommended.



*Figure 5. Case study 2 MNL model t-ratio results*



*Figure 6. Case study 2 MMNL model t-ratio results*

In addition to examining the *t*-ratios, Table 9 reports the MSE and RAE statistics for the MMNL model. Once more, for reasons of space we omit the results for the MNL model. For the MMNL model, we note a significant improvement in the MSE and RAE statistics once a sample size of 30 is achieved. Before this, despite producing decent *t*-ratios (see Figure 6), the parameter estimates appear to be significantly biased. Once 30 respondents are sampled however, the bias appears to largely disappear. As such, despite an *S*-error for some parameters of only five respondents, in this instance, at least 30 respondents should be sampled. Although not shown, similar results were found for the MNL model once 15 respondents were sampled.

**5.3 Case Study 3: Empirical Data Set**

The last case study utilises empirical data collected in the Netherlands in 2007 for a stated choice experiment involving an airline ticket choice context. A total of 204 respondents completed the experiment. As with the first two case studies, MNL and MMNL models are estimated on the data bootstrapped at different sample sizes. The experimental design used had 18 choice tasks with three blocks meaning each respondent saw only six. To preserve the design properties, the bootstrapping was performed in such a way that respondents were sampled equally from each block. Sample sizes divisible by 18 respondents were examined, starting with 18 respondents incrementally increased up until 180 respondents. The following base utility function was used for both the MNL and MMNL model runs. For the MMNL model, all parameters with the exception of the airline dummies were estimated as random parameters assuming a Multivariate Normal distribution.

*Table 9. Case Study 2 MSE and RAE results for the MMNL model*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *N* | β1 (mu) | β1 (sigma) | β2 | β3 | β4 (mu) | β4 (sigma) | β5 |
| *MSE* | | | | | | | |
| 5 | 311.462 | 60.614 | 2612.938 | 738.722 | 677.935 | 28.138 | 1056.629 |
| 10 | 155.538 | 26.209 | 719.109 | 245.895 | 199.390 | 16.876 | 376.511 |
| 15 | 60.772 | 30.918 | 321.574 | 59.085 | 106.401 | 12.780 | 88.760 |
| 20 | 23.902 | 13.168 | 115.672 | 41.158 | 27.390 | 7.077 | 3.092 |
| 25 | 7.015 | 6.912 | 77.299 | 31.386 | 25.504 | 2.569 | 23.707 |
| 30 | 0.083 | 0.022 | 0.679 | 0.167 | 0.151 | 0.017 | 0.370 |
| 35 | 0.066 | 0.019 | 0.479 | 0.124 | 0.096 | 0.012 | 0.226 |
| 40 | 0.053 | 0.015 | 0.445 | 0.140 | 0.082 | 0.012 | 0.190 |
| 45 | 0.045 | 0.013 | 0.380 | 0.104 | 0.080 | 0.012 | 0.144 |
| 50 | 0.036 | 0.010 | 0.302 | 0.081 | 0.056 | 0.011 | 0.140 |
| 55 | 0.039 | 0.010 | 0.296 | 0.077 | 0.054 | 0.010 | 0.132 |
| 60 | 0.033 | 0.008 | 0.257 | 0.064 | 0.044 | 0.007 | 0.133 |
| 65 | 0.029 | 0.008 | 0.237 | 0.067 | 0.041 | 0.006 | 0.110 |
| 70 | 0.033 | 0.008 | 0.194 | 0.050 | 0.042 | 0.006 | 0.094 |
| 75 | 0.030 | 0.006 | 0.211 | 0.046 | 0.037 | 0.006 | 0.084 |
| 80 | 0.030 | 0.007 | 0.206 | 0.054 | 0.035 | 0.006 | 0.079 |
| 85 | 0.026 | 0.006 | 0.215 | 0.054 | 0.031 | 0.006 | 0.079 |
| 90 | 0.024 | 0.006 | 0.174 | 0.040 | 0.027 | 0.005 | 0.077 |
| 95 | 0.021 | 0.004 | 0.155 | 0.035 | 0.025 | 0.005 | 0.064 |
| 100 | 0.021 | 0.004 | 0.147 | 0.034 | 0.025 | 0.005 | 0.055 |
| 105 | 0.017 | 0.007 | 0.094 | 0.028 | 0.014 | 0.004 | 0.044 |
| 110 | 0.020 | 0.004 | 0.124 | 0.030 | 0.019 | 0.005 | 0.050 |
| 115 | 0.015 | 0.005 | 0.089 | 0.025 | 0.012 | 0.003 | 0.043 |
| 120 | 0.013 | 0.003 | 0.081 | 0.026 | 0.009 | 0.003 | 0.045 |

In generating the original design, a different utility specification was employed to that given in Equation (8). Further, the design was optimised assuming an MNL model specification. Finally, the model results estimated on the full data suggest different population level parameters than were used in the design generation process (see Bliemer, *et al*, 2009b for more details on the design and models estimated on the full data set). As such, the results presented here represent an examination of the issues of sample size given model, utility specification and prior parameter misspecification.

*Table 9. Continued*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *N* | β1 (mu) | β1 (sigma) | β2 | β3 | β4 (mu) | β4 (sigma) | β5 |
| *RAE* | | | | | | | |
| 5 | 15.609 | 13.413 | 21.155 | 20.558 | 18.682 | 12.936 | 31.529 |
| 10 | 5.801 | 5.484 | 6.902 | 7.197 | 5.959 | 6.128 | 10.627 |
| 15 | 3.533 | 5.774 | 3.921 | 3.096 | 3.657 | 5.460 | 4.574 |
| 20 | 1.540 | 2.556 | 1.704 | 1.829 | 1.359 | 2.913 | 1.342 |
| 25 | 0.824 | 1.566 | 1.265 | 1.382 | 1.112 | 1.717 | 1.851 |
| 30 | 0.294 | 0.381 | 0.407 | 0.378 | 0.294 | 0.525 | 0.744 |
| 35 | 0.261 | 0.349 | 0.356 | 0.347 | 0.251 | 0.440 | 0.631 |
| 40 | 0.248 | 0.312 | 0.338 | 0.360 | 0.227 | 0.446 | 0.571 |
| 45 | 0.239 | 0.303 | 0.318 | 0.315 | 0.227 | 0.439 | 0.498 |
| 50 | 0.213 | 0.262 | 0.288 | 0.281 | 0.209 | 0.387 | 0.501 |
| 55 | 0.205 | 0.259 | 0.268 | 0.263 | 0.186 | 0.374 | 0.468 |
| 60 | 0.199 | 0.221 | 0.263 | 0.258 | 0.173 | 0.337 | 0.483 |
| 65 | 0.189 | 0.233 | 0.246 | 0.256 | 0.168 | 0.297 | 0.438 |
| 70 | 0.187 | 0.224 | 0.221 | 0.215 | 0.173 | 0.298 | 0.409 |
| 75 | 0.188 | 0.221 | 0.233 | 0.205 | 0.156 | 0.302 | 0.393 |
| 80 | 0.177 | 0.217 | 0.232 | 0.228 | 0.164 | 0.298 | 0.380 |
| 85 | 0.184 | 0.208 | 0.224 | 0.220 | 0.163 | 0.286 | 0.390 |
| 90 | 0.170 | 0.202 | 0.223 | 0.197 | 0.140 | 0.280 | 0.366 |
| 95 | 0.162 | 0.162 | 0.205 | 0.177 | 0.139 | 0.259 | 0.336 |
| 100 | 0.164 | 0.172 | 0.206 | 0.182 | 0.130 | 0.265 | 0.309 |
| 105 | 0.145 | 0.207 | 0.163 | 0.159 | 0.104 | 0.241 | 0.267 |
| 110 | 0.152 | 0.168 | 0.182 | 0.170 | 0.117 | 0.254 | 0.283 |
| 115 | 0.143 | 0.176 | 0.163 | 0.161 | 0.095 | 0.208 | 0.287 |
| 120 | 0.131 | 0.152 | 0.147 | 0.157 | 0.075 | 0.203 | 0.286 |

|  |  |
| --- | --- |
|  | (8) |

Given the use of empirical data, the true population parameters are not known. To examine how well the *S*-error estimates perform in predicting the sample size requirements for the design, we first estimate a model on the full data (*N* = 204) and assume these to be the true population parameters. These parameter estimates are then used to estimate the *S*-errors for the design as well as the predicted *t*-ratios at different sample sizes. Next, a series of bootstrap runs are performed using the same sample sizes used in the previous step. The average *t*-ratios over the bootstrap runs for each of the parameters are then calculated at the different sample sizes. The predicted *t*-ratios based on the design and assumed population parameter estimates can then be compared to the observed bootstrapped *t*-ratios. Figure 7 plots the predicted *t*-ratios against the observed bootstrapped *t*-ratios for the MNL and MMNL logit model specifications at the various sample sizes examined. In the figure, each data point shown depicts a different parameter *t*-ratio.

For the MNL model, the overall prediction for the *t*-ratios is quite good, even at small sample sizes. Indeed, the method is able to predict quite well which *t*-ratios will be observed to be low and which *t*-ratios will be observed to be high. A similar pattern emerges for the MMNL model. For both model types, we note that the observed *t*-ratios are smaller than the predicted *t*-ratios. This once more confirms that the predicted *t*-ratios can be seen as a lower bound, where for the MNL model, the bound is general quite good predictor, although some outliers do exist. The results for the MMNL model also support the notion that the predicted *t*-ratios represent a lower bound however the bound provides a less tight predictor of the actual observed value. This however is not surprising given that the MMNL model represents a misspecification of not only the model type assumed during the design generation phase, but also misspecification of the utility function and prior parameter estimates.

Based on the results for the MNL and MMNL models, we are able to conclude that the theoretical design principles used to calculate the *S*-error and predict what *t*-ratios will likely be observed do provide useful and in the case of the MNL model, quite accurate predictions of the *t*-ratios prior to conducting the survey. This suggests that it should be possible to identify parameters in advance that will likely not be statistically significant in estimation, or at a minimum, identify which parameters are most likely to cause issues in terms of obtaining statistical significance. Further, the *S*-error calculations proposed by Bliemer and Rose (2005; 2009; 2010) and Rose and Bliemer (2005) do appear to provide adequate information as to the required theoretical minimum sample size requirements of the experiment, however as can be seen based on the results of the MMNL model, larger sample sizes than those suggested by the measure should always be captured. In this way, the analyst will be protected in terms of the various sources of possible misspecifications that might occur during the study.

1. **Discussion and Conclusion**

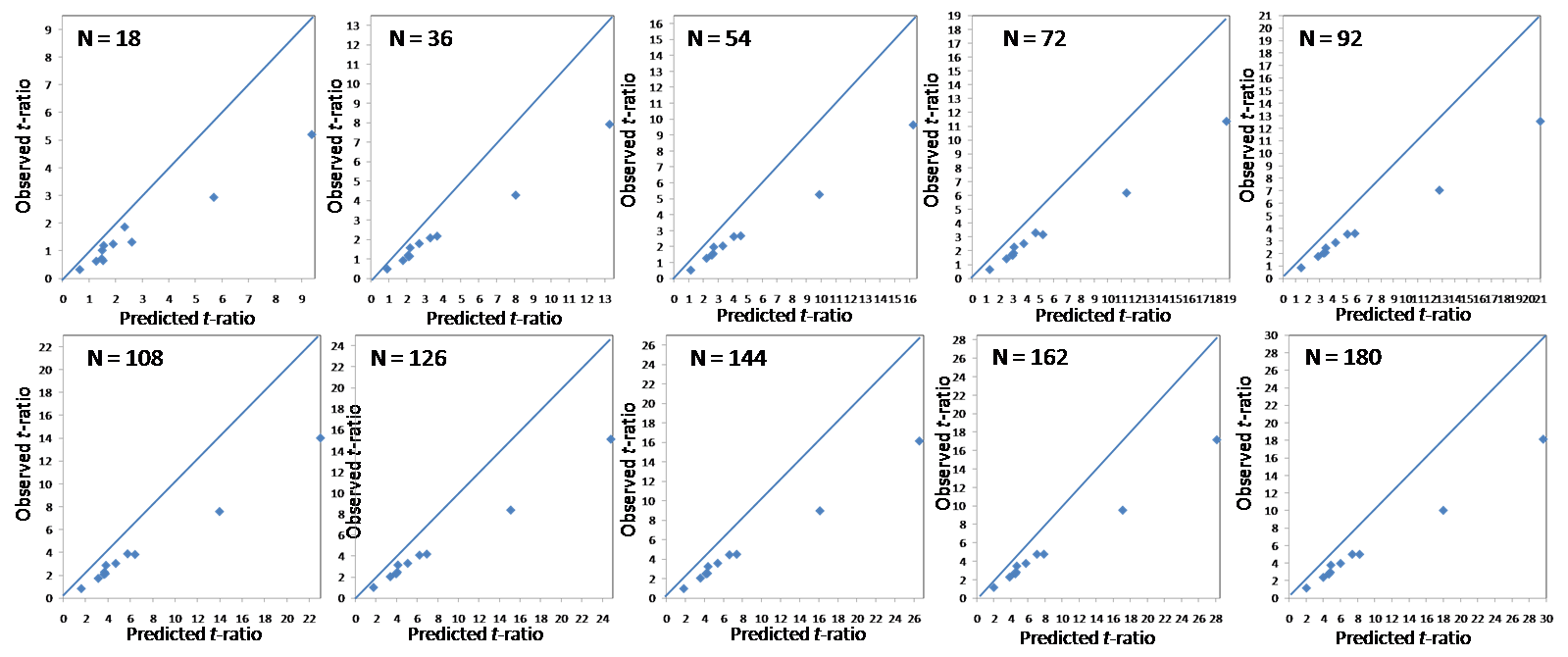
This paper has, using three different case studies, two based on simulated data and one using empirical data, systematically examined issues related to the sample size requirements for discrete choice experiments. The paper outlined the *S*-error sample size calculation proposed by Bliemer and Rose (2005; 2009; 2010) and Rose and Bliemer (2005) and explored the relationship of the analytically derived measure to estimated model results. In doing so, several different design types and several different econometric model types were examined. All three case studies also explored issues related to various possible experimental design misspecifications and how these might impact upon the sample size requirements for choice experiments.

This study has shown that outputs derived from the expected (asymptotic) variance-covariance matrices computed during the design generation process may be used to predict the outcomes of models estimated post data collection. In particular, the paper has shown that the theoretical minimum sample size requirements can be computed (for a given preferred minimum *t*-ratio) and that *t­*-ratios can be predicted with a good degree of accuracy before the experiment goes to field. Nevertheless, consistent with the earlier work of Bliemer and Rose (2005; 2009; 2010) and Rose and Bliemer (2005), these predictions represent theoretical minimum bounds and larger sample sizes are recommended. Indeed, based on the case study, results presented, sample sizes three to four times that suggested might be required to overcome different possible sources of misspecification that might occur during the study.

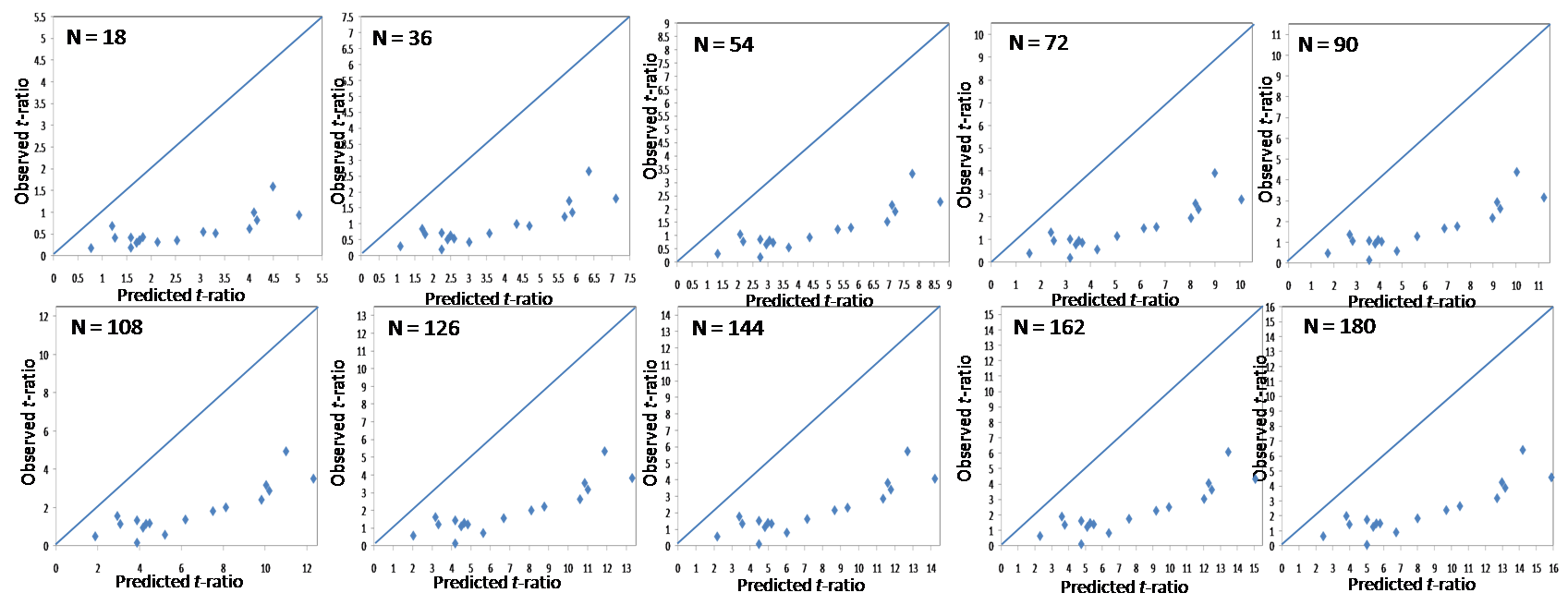
At a minimum however, the results shown herein do provide evidence as to what parameters will likely experience difficulties in retrieving statistically significant *t­*-ratios. Armed with such knowledge, the analyst may be able to take some form of preventative action prior to going to field. For example, the analyst may seek to undertake further qualitative research to seek an alternative representation for the attribute or attributes in question or conduct an additional pilot study to determine whether the priors assumed in the design generation can be improved upon, hence leading to a better design. Alternatively, such knowledge may allow a rethink as to the attribute levels used, for example resulting in a changing of the attribute level range used for quantitative attributes, or changing the number of levels used for either qualitative or quantitative attributes. Such remedies can be explored prior to data collection and possibly result in considerable savings being achieved, both financial and time wise.

In addition to examining the *t*-ratios, this paper also sought to examine at what sample size the parameter estimates become stable. More often than not, the *S*-error sample size calculation suggests (at a *t*-ratio of 1.96) sample sizes as low as four or five respondents with occasionally lower sample sizes being suggested. At such small sample sizes, the question becomes whether the parameter estimates are stable enough such that adding an additional respondent will change the results dramatically. To examine this, we have produced MSE and RAE statistics for the first two case studies. Both statistics measure bias of estimates from a known value, in this case, the estimated parameters from the known population parameters. In the first case study, the smallest sample size used was 25 respondents. Independent of the model used, the MSE and RAE statistics suggest that even with only 25 respondents, this sample size is sufficient to reproduce parameter estimates which are close to the true population parameters. In the second case study, sample sizes as small as five respondents were examined. In this case, the MSE and RAE statistics suggest that for the MNL model, at least 15 respondents are required whilst for the MMNL model, a minimum of 30 respondents are required. Not surprising, both case studies suggest that larger sample sizes however reduce the risk of obtaining biased (from the known parameter values) parameters. Given the findings of the second case study, we recommend that at a minimum, at least 30 respondents be sampled for any discrete choice study. Nevertheless, despite this recommendation, the methods used here may be replicated for any other study, and we further recommend that in addition to using the *S*-error sample size calculation, that those generating designs use Monte Carlo simulations to test the impact of the final design at different sample sizes prior to going to field.

In conclusion, our research supports the claims made by those researching in the area of efficient designs and sample size for discrete choice studies. Nevertheless, we encourage further studies, similar to the one described here, encompassing greater numbers of types of designs and models be carried out. Indeed, further empirical research is urgently needed to understand whether the predominately simulated results undertaken in research into the generation of experimental designs for stated choice translate into reality.



*Figure 4a. Case study 3 MNL model results predicted versus observed t-ratios*



*Figure 4b. Case study 3 MMNL model results predicted versus observed t-ratios*

**References**

Anderson, D.A. and Wiley, J.B., 1992. Efficient choice set designs for estimating cross effect models. *Marketing Letters*, 3(4), pp.357-370.

Bliemer, M.C.J. and Rose, J.M., 2005. Efficiency and sample size requirements for stated choice studies. Working paper: ITLS-WP-05-08. Sydney: Institute of Transport Studies and Logistics.

Bliemer, M.C.J. and Rose, J.M., 2006. Designing stated choice experiments: the state of the art. In: *11th International Conference on Travel Behaviour Research*. Kyoto, Japan, 16-20 August 2006.

Bliemer, M.C.J. and Rose, J.M., 2009. Efficiency and sample size requirements for stated choice studies. In: *88th Annual Meeting of the Transportation Research Board*. Washington DC, USA, 11-15 January 2009.

Bliemer, M.C.J. and Rose, J.M., 2010. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B*, 44(6), pp.720-734.

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

Bliemer, M.C.J., Rose, J.M. and Beelaerts van Blokland, R., 2009b. Experimental design influences on stated choice outputs. In: *European Transport Conference 2009*. Leeuwenhorst, The Netherlands, 5-7 October 2009.

Bliemer, M.C.J., Rose, J.M. and Hess, S., 2008. Approximation of Bayesian efficiency in experiment choice designs. *Journal of Choice Modelling*, 1(1), pp.98-127.

Bunch, D.S., Louviere, J.J., and Anderson D.A., 1996. A comparison of experimental design strategies for choice-based conjoint analysis with generic-attribute multinomial logit models. Working Paper. Davis: Graduate School of Management, University of California, Davis.

Burgess, L., 2007. *Discrete Choice Experiments [computer software]*. [online] Available at <http://crsu.science.uts.edu.au/choice/>.

Burgess, L. and Street, D.J., 2005. Optimal designs for choice experiments with asymmetric attributes. *Journal of Statistical Planning and Inference*,134(1), pp.288-301.

Carlsson, F. and Martinsson, P., 2002. Design techniques for stated preference methods in health economics. *Health Economics*,12(4), pp.281-294.

Cirillo, C., Daly, A.J. and Lindveld, K., 2000. Eliminating Bias Bue to the Repeated Measurements Problem in SP Data. In: J.de D. Ortúzar, ed. *Stated Preference Modelling Techniques: PTRC Perspectives 4*. London: PTRC Education and Research Services.

Fowkes, T. and Wardman, M.R., 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), pp.27-44.

Hedayat, A.S., Sloane, N.J.A. and Stufken, J., 1999. *Orthogonal Arrays: Theory and Applications.* New York: Springer-Verlag.

Hess, S. and Rose, J.M., 2009. Allowing for intra-respondent variations in coefficients estimated on stated preference data. *Transportation Research Part B*, 43(6), pp.708-719.

Huber, J. and Zwerina, K., 1996. The importance of utility balance and efficient choice designs. *Journal of Marketing Research*, 33(2), pp.307-317.

Kanninen, B.J., 2002. Optimal design for multinomial choice experiments. *Journal of Marketing Research*, 39(12), pp.214-217.

Kessels, R., Goos, P. and Vandebroek, M., 2006. A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*, 43(3), pp.409-419.

Lazari, A.G. and Anderson, D.A., 1994. Designs of discrete choice experiments for estimating both attribute and availability cross effects. *Journal of Marketing Research*, 31(3), pp.375-383.

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

McFadden, D., 1974. Conditional Logit Analysis of Qualitative Choice Behaviour. In: P. Zarembka, ed. *Frontiers in Econometrics*. New York: Academic Press, pp.105-142.

Ortúzar, J.de D. and Willumsen, L.G., 2001. *Modelling Transport*. 3rd ed. Chichester: John Wiley and Sons.

Ortúzar, J.de D., Roncagliolo, D.A and Velarde, U.C., 2000. Interactions and Independence in Stated Preference Modelling. In: J.de D. Ortúzar, ed. *Stated Preference Modelling Techniques: PTRC Perspectives 4*. London: PTRC Education and Research Services.

Revelt, D. and Train, K., 1998. Mixed logit with repeated choices: households’ choices of appliance efficiency level. *Review of Economics and Statistics*, 80(4), pp.647-657.

Rose, J.M. and Bliemer, M.C.J., 2005. Sample optimality in the design of stated choice experiments. Working paper: ITLS-WP-05-13. Sydney: Institute of Transport Studies and Logistics.

Rose, J.M. and Bliemer, M.C.J., 2008. Stated Preference Experimental Design Strategies. In: D.A. Hensher and K.J. Button, eds. *Handbook of Transport Modelling*.Oxford: Elsevier, pp.151-180.

Rose, J.M., Scarpa, R. and Bliemer, M.C.J., 2009. Incorporating model uncertainty into the generation of efficient stated choice experiments: a model averaging approach*.* In: *International Choice Modelling Conference 2009*. Yorkshire, UK, March30-April 1 2009*.*

Sándor, Z. and Wedel, M., 2001. Designing conjoint choice experiments using managers’ prior beliefs. *Journal of Marketing Research*, 38(4), pp.430-444.

Sándor, Z. and Wedel, M., 2002. Profile construction in experimental choice designs for mixed logit models. *Marketing Science*, 21(4), pp.455-475.

Sándor, Z. and Wedel, M., 2005. Heterogeneous conjoint choice designs. *Journal of Marketing Research*, 42(2), pp.210-218.

Street, D.J. and Burgess, L., 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), pp.185-199.

Street, D.J. and Burgess, L., 2007. *The Construction of Optimal Stated Choice Experiments: Theory and Methods*. Hoboken: Wiley.

Street, D.J., Burgess, L., and Louviere, J.J., 2005. Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments. *International Journal of Research in Marketing*, 22(4), pp.459-470.

Train, K., 2003. *Discrete Choice Methods with Simulation.* Cambridge: Cambridge University Press.

**Appendix**

*Table 10. Case Study 3 design*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **s** | **j** | **A** | **PR** | **DT** | **TR** | **EP** | **ET** | **s** | **j** | **A** | **PR** | **DT** | **TR** | **EP** | **ET** |
| 1 | 1 | 2 | 100 | 12 | 60 | 3 | 30 | 3 | 4 | 5 | 75 | 6 | 60 | 12 | 80 |
| 1 | 2 | 1 | 50 | 18 | 0 | 15 | 60 | 3 | 5 | 2 | 50 | 18 | 120 | 12 | 60 |
| 1 | 3 | 2 | 100 | 18 | 60 | 1 | 40 | 4 | 1 | 2 | 100 | 18 | 60 | 12 | 100 |
| 1 | 4 | 4 | 75 | 12 | 120 | 1 | 30 | 4 | 2 | 3 | 50 | 18 | 120 | 5 | 20 |
| 1 | 5 | 0 | 50 | 6 | 60 | 1 | 30 | 4 | 3 | 1 | 50 | 6 | 0 | 9 | 100 |
| 2 | 1 | 5 | 75 | 12 | 60 | 5 | 20 | 4 | 4 | 5 | 50 | 6 | 120 | 9 | 80 |
| 2 | 2 | 2 | 100 | 18 | 60 | 3 | 40 | 4 | 5 | 4 | 75 | 12 | 0 | 15 | 60 |
| 2 | 3 | 4 | 75 | 18 | 0 | 12 | 60 | 5 | 1 | 1 | 50 | 6 | 0 | 1 | 30 |
| 2 | 4 | 3 | 100 | 6 | 60 | 3 | 40 | 5 | 2 | 4 | 50 | 12 | 120 | 15 | 100 |
| 2 | 5 | 3 | 50 | 6 | 120 | 9 | 100 | 5 | 3 | 5 | 50 | 18 | 60 | 15 | 60 |
| 3 | 1 | 0 | 75 | 12 | 0 | 9 | 100 | 5 | 4 | 0 | 50 | 18 | 120 | 15 | 80 |
| 3 | 2 | 4 | 75 | 6 | 0 | 5 | 40 | 5 | 5 | 4 | 100 | 18 | 0 | 1 | 20 |
| 3 | 3 | 3 | 75 | 12 | 120 | 9 | 60 | 6 | 1 | 3 | 50 | 6 | 60 | 15 | 80 |

*Table 10. Continued*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **s** | **j** | **A** | **PR** | **DT** | **TR** | **EP** | **ET** | **s** | **j** | **A** | **PR** | **DT** | **TR** | **EP** | **ET** |
| 6 | 2 | 0 | 50 | 12 | 60 | 9 | 60 | 12 | 4 | 5 | 50 | 18 | 0 | 12 | 100 |
| 6 | 3 | 5 | 75 | 12 | 120 | 3 | 30 | 12 | 5 | 1 | 100 | 18 | 60 | 3 | 30 |
| 6 | 4 | 2 | 100 | 18 | 0 | 5 | 30 | 13 | 1 | 3 | 75 | 6 | 0 | 12 | 60 |
| 6 | 5 | 2 | 50 | 6 | 120 | 5 | 40 | 13 | 2 | 5 | 75 | 18 | 120 | 1 | 30 |
| 7 | 1 | 4 | 50 | 18 | 120 | 1 | 40 | 13 | 3 | 0 | 50 | 12 | 60 | 12 | 100 |
| 7 | 2 | 0 | 100 | 6 | 60 | 3 | 30 | 13 | 4 | 1 | 100 | 6 | 60 | 3 | 30 |
| 7 | 3 | 1 | 75 | 6 | 60 | 15 | 80 | 13 | 5 | 2 | 100 | 18 | 0 | 3 | 20 |
| 7 | 4 | 2 | 100 | 12 | 0 | 3 | 20 | 14 | 1 | 2 | 75 | 12 | 120 | 5 | 40 |
| 7 | 5 | 5 | 100 | 12 | 60 | 12 | 100 | 14 | 2 | 5 | 50 | 6 | 0 | 15 | 80 |
| 8 | 1 | 1 | 100 | 18 | 0 | 3 | 20 | 14 | 3 | 4 | 100 | 12 | 120 | 3 | 20 |
| 8 | 2 | 2 | 50 | 12 | 120 | 9 | 100 | 14 | 4 | 3 | 50 | 18 | 0 | 9 | 100 |
| 8 | 3 | 3 | 100 | 6 | 120 | 5 | 20 | 14 | 5 | 1 | 100 | 18 | 60 | 12 | 80 |
| 8 | 4 | 0 | 100 | 12 | 0 | 15 | 60 | 15 | 1 | 5 | 100 | 12 | 60 | 9 | 60 |
| 8 | 5 | 3 | 100 | 18 | 120 | 5 | 20 | 15 | 2 | 3 | 75 | 6 | 0 | 9 | 80 |
| 9 | 1 | 5 | 50 | 18 | 120 | 12 | 80 | 15 | 3 | 4 | 50 | 18 | 0 | 15 | 100 |
| 9 | 2 | 4 | 100 | 6 | 60 | 1 | 20 | 15 | 4 | 1 | 100 | 12 | 60 | 1 | 20 |
| 9 | 3 | 0 | 75 | 12 | 120 | 5 | 40 | 15 | 5 | 0 | 75 | 6 | 120 | 5 | 30 |
| 9 | 4 | 4 | 75 | 18 | 0 | 1 | 40 | 16 | 1 | 4 | 75 | 6 | 0 | 3 | 40 |
| 9 | 5 | 1 | 75 | 12 | 0 | 15 | 80 | 16 | 2 | 1 | 75 | 18 | 0 | 12 | 80 |
| 10 | 1 | 0 | 100 | 6 | 120 | 9 | 80 | 16 | 3 | 2 | 75 | 18 | 60 | 1 | 20 |
| 10 | 2 | 3 | 100 | 18 | 0 | 3 | 20 | 16 | 4 | 3 | 50 | 12 | 120 | 5 | 20 |
| 10 | 3 | 1 | 50 | 18 | 60 | 12 | 80 | 16 | 5 | 5 | 75 | 6 | 60 | 1 | 40 |
| 10 | 4 | 0 | 50 | 6 | 120 | 9 | 60 | 17 | 1 | 3 | 100 | 12 | 60 | 15 | 100 |
| 10 | 5 | 5 | 75 | 12 | 0 | 3 | 40 | 17 | 2 | 1 | 75 | 12 | 120 | 1 | 40 |
| 11 | 1 | 1 | 50 | 18 | 120 | 1 | 20 | 17 | 3 | 2 | 50 | 6 | 0 | 9 | 80 |
| 11 | 2 | 0 | 100 | 12 | 120 | 5 | 30 | 17 | 4 | 4 | 75 | 18 | 60 | 12 | 60 |
| 11 | 3 | 5 | 100 | 6 | 0 | 5 | 30 | 17 | 5 | 0 | 75 | 12 | 60 | 15 | 100 |
| 11 | 4 | 2 | 75 | 6 | 60 | 5 | 40 | 18 | 1 | 0 | 75 | 18 | 0 | 5 | 30 |
| 11 | 5 | 3 | 50 | 12 | 0 | 9 | 80 | 18 | 2 | 5 | 75 | 12 | 60 | 12 | 100 |
| 12 | 1 | 4 | 50 | 6 | 120 | 15 | 60 | 18 | 3 | 3 | 100 | 6 | 120 | 3 | 40 |
| 12 | 2 | 2 | 100 | 6 | 60 | 12 | 60 | 18 | 4 | 1 | 75 | 12 | 120 | 15 | 100 |
| 12 | 3 | 0 | 100 | 12 | 0 | 1 | 30 | 18 | 5 | 4 | 50 | 6 | 120 | 9 | 60 |

1. http://crsu.science.uts.edu.au/choice/choice.html [↑](#footnote-ref-1)
2. We use Halton sequences to generating the sample parameters as well as the EV1 draws. [↑](#footnote-ref-2)