EXPERIMENTAL DESIGN INFLUENCES ON STATED CHOICE OUTPUTS: AN EMPIRICAL STUDY IN AIR TRAVEL CHOICE

Michiel C.J. Bliemer Delft University of Technology, The University of Sydney, Goudappel Coffeng John M. Rose The University of Sydney Reinier E.F. Beelaerts van Blokland Delft University of Technology

1. INTRODUCTION

Discrete choice experiments (DCE) have grown to become the primary source of data for obtaining estimates of behavioural importance such as consumer preferences for various transport goods and services or willingness to pay (WTP) measures for specific attributes such as travel time savings. Over time, as the discrete choice modelling literature has matured, a number of econometrically more advanced models able to uncover an increasing degree of behavioural richness have been developed, typified by the rapid progression from multinomial logit (MNL), nested logit (NL), cross-correlated NL and mixed multinomial logit (MMNL) models. At the same time, advancements in the construction of experimental designs that underlie DCE have been limited and somewhat more erratic in terms of acceptance within the wider literature. This is not to suggest however that advancements have not been made.

Unfortunately, whilst information related to which alternatives, attributes and attribute levels to use may come from secondary data sources or qualitative research such as focus groups and in-depth interviews, the precise method used to construct the underlying experimental design remain solely at the discretion of the researcher. Whilst there exist multiple possible construction methods, each of which make different assumptions during the design generation process, unlike estimation problems where a number of texts describe the advantages and disadvantages of various models (see e.g., Train 2003), there regrettably exists little guidance as to which particular method to select when generating an experimental design for DCE type studies. A poor choice of experimental design, or one based on an inadequate or incorrect set of assumptions, may result in poor data quality. In turn, poor quality data may result in erroneous conclusions being reached, or at the very least, less reliable parameter estimates at a given sample size. At issue however, is that the advancements that have been made in relation to experimental design construction for DCEs have largely been theoretical in nature, with little empirical work demonstrating whether the theoretical advantages of the various identified design strategies actually translate into practice.

As such, the most widely used type of design for DCEs remain what we term orthogonal fractional designs or simply orthogonal designs (see Section 2.1). The orthogonality (or otherwise) of an experimental design relates to the correlation structure between the attributes of the design with designs in which all between attribute correlations are zero being said to be orthogonal (in some cases, this definition of an orthogonal design may be relaxed to define orthogonality as occurring when all attribute correlations are zero within alternatives but not necessarily between alternatives; see Louviere et al. (2000) for discussion on sequential versus simultaneous generation of orthogonal designs). The popular use of orthogonal designs stems firstly from historical impetus, and secondly from a lack of guidance within the literature as to whether such designs represent best practice. In relation to historical impetus, experimental design theory as related to early DCE was simply borrowed from existing experimental design theory dealing primarily with linear models, in particular, ANOVA and regression type models (see e.g., Louviere et al. 1983). This resulted in the large scale use of orthogonal designs which continued unquestioned for many years to come. The lack of empirical evidence for or against the use of such designs stems from the typical use within study of a single experimental design. The failure to vary different experimental designs, whilst understandable given that most empirical studies are focused on some particular practical goal other than testing the impact of different experimental designs, makes comparisons of different design construction techniques impossible, as well as preventing the offering of a robust set of guidelines for best practice in generating experimental designs specifically for DCEs. This coupled with the fact that orthogonal designs appear to have worked well over the years has also resulted in their widespread use for DCE studies.

Nevertheless, alternatives to orthogonal designs have been available for some time. Beginning with Fowkes et al. (1993), Bunch et al. (1996) and Huber and Zwerina (1996), new experimental design theories began to emerge in the early to mid 1990s specifically for the generation of experiments for the non-linear discrete choice models often associated with DCE. Further theoretical advancements in the field of experimental designs have been made since including, but not limited to work undertaken by Bliemer et al. (2009). Burgess and Street (2003). Ferrini and Scarpa (2006), Kanninen (2002), Kessels et al. (2006), Sándor and Wedel (2001, 2002, 2005), Rose and Bliemer (2008), Scarpa and Rose (2008), Street and Burgess (2004) and Toner et al. (1999). The types of designs constructed by these researchers, typically referred to as efficient designs, each have the common goal of seeking to minimize the determinant of the asymptotic variance-covariance (AVC) matrix of models estimated on data collected using the designs, which essentially minimizes the standard errors. Differences in the construction methods posed by each of the above researchers however lie solely in the assumptions made about how to a priori estimate what the AVC matrix of a design might look like without having first collected any data.

The objective of this paper is not to detail the advancements in generating experimental designs for DCEs. This has been done elsewhere (see e.g., Bliemer and Rose 2006, Rose and Bliemer *in press*, 2008). Rather, the objective of this paper is to empirically compare the results obtained from the two different experimental designs types that have been proposed and mainly used within the literature. In this respect, this paper joins a very small but growing area of research that seeks to clarify whether the theoretical advancements claimed by researchers in the area of experimental design generation for DCEs translates into practice. It is hoped that such research will ultimately produce practical guidelines for the construction of DCE experimental designs.

The remainder of the paper is organised as follows. In the next section, we provide a brief literature review of several of the top tier transportation journals, in terms of the types of experimental designs that have been used in the past. In this section, we also provide a review of the literature exploring the empirical impacts of different experimental design types. The following section details the experimental context, designs and data used for empirical comparison within this paper. Section 4 provides a discussion of the results before concluding comments are provided.

2. LITERATURE REVIEW

In this section we report the results of literature review of a subset of tier 1 transportation journals specifically looking at the types of experimental designs used. We then go onto to discuss the literature examining empirically the impacts of various experimental design types have upon behavioural outcomes.

2.1 A Review of the Transportation Literature

Table 1 summarizes the literature using DCE methods published in a subset of tier one transportation journals during the period of January 2000 to August 2009¹. Only articles in which the type of experimental design used could be determined are listed (i.e., articles such as Hess et al. (2007) which report the use of DCEs but do not provide any details as to the experimental designs used are excluded from the table). A total of 64 research papers were identified yielding 61 unique DCE experimental designs. Of these 61 studies, 40 (66 percent) utilized an orthogonal design, 12 a D-efficient designs (20 percent), seven (11 percent) randomly assigned attribute levels shown to respondents and three (3 percent) used an adaptive design approach, alternating the levels shown to respondents based on the respondents previous answers.

The number of alternatives, J, for each study reported range between two and six and several of these studies included a current or reference alternative from which the DCE alternatives are pivoted from (e.g., Hensher and Prioni 2002). The number of attributes, K, per alternative varied from three to 46 and the number of levels per attribute from two to 30. Some studies, such as Leitham et al. (2000), employed designs in which different attributes of the experiment had different numbers of attribute levels. Also reported, when detailed in the original study, are the fractional factorial dimensions and the size of the experimental designs used. The size of the design represents the total number of choice sets, S, generated. The majority of studies reviewed employed only a subset of the total possible number of choice sets. In determining which choice sets to give to which respondents, most studies used an additional blocking column in the design (39, or 64 percent of designs), however, some randomly assigned choice sets to respondents (8, or 13 percent). Three (5 percent) studies gave each respondent all choice sets that were generated as part of the design. It was not possible to determine how the choice tasks were allocated to respondents in 11 (18 percent) of the studies reviewed. The number of choice sets given to any particular respondent is shown in the eighth column of Table 1.

The number of respondents, *N*, surveyed is shown in the third last column of Table 1**Error! Reference source not found.**² The sample sizes reported represent the number of complete surveys returned that were deemed useable for each study. Related to the sample size is the number of observations, *N.S*, defined here as the number of choice sets used in the model estimation process. The number of observations ranged from a low of 60 (Brewer and Hensher 2000) to a high of 46,632 (Teichert et al. 2008). The median number of observations over all the studies in which the number of observations in the study can be readily determined³ is 2,688.

Reference	Design Type	Labelled/ Unlabeled	S assignment	J	ĸ	S	Shown	Design size	Ν	N.S.	Models Estimated
Brewer, A.M. & Hensher, D.A. (2000)	Orthogonal	Unlabeled	Random	3	10	27	3	3 ¹⁰	20	60	MNL
Brownstone, D. Bunch, D.S. & Train, K. (2000)	Random	Unlabeled	N/S	3	14	N/S	1	4 ¹⁴	7387	4747	ML model
Leitham, S. McQuaid, R.W. & Nelson, J.D. (2000)	Orthogonal	Unlabeled	Random	2	7	16	2	2 ⁴ ×3 ³	40	640	MNL
Ortúzar, JdeD. Iacobelli A, & Valeze, C. (2000)	Orthogonal	Labelled	All	2	7	8	8	2 ⁷	357	2774	Logit
Wang, D. Borgers, A. Oppewal, H. & Timmermans, H. (2000)	Orthogonal	Unlabeled	Blocked	2	14	81	9	2 ³ ×3 ¹¹	335	3015	MNL
Hensher, D.A. (2001)	Orthogonal	Unlabeled	Blocked	2+rp	6	64	16	4 ² ×6 ¹	114	1824	MMNL
Hensher, D.A. (2001)	Orthogonal	Unlabeled	Blocked	2+rp	6	64	16	4 ¹²	198	3168	MNL/MMNL
Hensher, D.A. & King, J. (2001)	Orthogonal	Labelled	N/S	5	9	27	3	3 ⁹	536	1608	NL
Saelensminde, K. (2001)	Orthogonal	Unlabeled	Random	2	3	16	9	4 ³	508	4572	Logit scaling
Bhat, C.R. & Castelar, S. (2002)	Orthogonal	Labelled	Blocked	6	4	32	8	N/S	136	1068	MNL/NL
Cherchi, E. & Ortuzar, JdeD. (2002)	Orthogonal	Labelled	Blocked	3	4	27	8-9	3 ⁴	524	1396	NL (SP-RP)
Garrod, G.D., Scarpa, R. & Willis, K.G. (2002)	Orthogonal	Unlabeled	Random	2+no	5	72	8	2 ³ ×3 ²	414	3312	MNL
Hensher, D.A. & Greene, W.H. (2002)	Orthogonal	Labelled	N/S	5	3	N/S	4	3 ¹⁵	210	840	NL
Hensher, D.A. & Prioni, P. (2002)	Orthogonal	Unlabeled	Blocked	2+rp	13	81	3	3 ¹³	N/S	3489	MNL
Wang, B. Hensher, D.A. & Ton, T. (2002)	Orthogonal	Unlabeled	NA	1	9	27	27	3 ⁵ ×3 ⁴	194	5238	Ordered Probit
Greene, W.H. & Hensher, D.A. (2003)	Orthogonal	Labelled	Blocked	4	6	32	16	4 ⁶	274	4384	MNL/LC/MMNL
Hensher, D.A. & Greene, W.H. (2003)	Orthogonal	a: Unlabeled	Blocked	3+rp	6	32	16	4 ⁶	274	4384	MMNL
-	Orthogonal	b: Unlabeled	Blocked	2+rp	6	64	16	4 ¹²	143	2288	MMNL
-	Orthogonal	c: Unlabeled	Blocked	2+rp	8	N/S	16	4 ⁵ ×2 ³	60	960	MMNL
Jovicic G. & Hansen, CO. (2003)	Orthogonal	Unlabeled	N/S	2	3-5	N/S	N/S	N/S	418/589/453	19,989	NL (SP-RP)
Hensher, D.A. (2004)	D-efficient	Unlabeled	Blocked	3-5	3-6	12-30	6-15	Varying	427	4593	MNL/HEV
Zhang, J. Timmermans, H., Borgers, A. & Wang, D. (2004)	Orthogonal	Unlabeled	Blocked	2	14	81	9	3 ¹⁴	1400	3015	MNL
Cantillo, V. & Ortuzar, JdeD. (2005)	Orthogonal	Unlabeled	Blocked	2	3	27	9	3 ³	342	3071	MNL/MMNL
Caussade, S., Ortuzar, JdeD., Rizzi, L.I. & Hensher, D.A. (2005)	D-efficient	Unlabeled	Blocked	3-5	3-6	12-30	6-15	Varying	403	8020	MNL/Hetero L
de Palma, A. & Picard, N. (2005)	Random	Unlabeled	N/A	N/S	N/S	N/S	N/S	N/S	4137	N/S	Ordered Probit
Saleh, W. & Farrell, S. (2005)	Orthogonal	Labelled	Random	3	8	27	7	3 ⁸	632	4424	MNL
Anderson, C.M., Das, C. & Tyrrel, T.J. (2006)	D-efficient	Unlabeled	Blocked	2	7	35	7	3 ⁵ ×4 ²	298	2063	MMNL
Bastin, F., Cirillo, C. & Toint, P.L. (2006)	Orthogonal	Labelled	N/S	4	7	N/S	3	3 ⁶ ×2 ¹	871	2602	MMNL
Bhat, C.R. & Sardesai, R. (2006)	Orthogonal	Labelled	Blocked	3⁄4	5 from 14	64	4	2 ¹¹ ×4 ³	319	1276	MMNL
Cantillo, V. Heydecker, B. & Ortuzar, JdeD. (2006)	D-efficient	Unlabeled	Blocked	3/5	3	N/S	N/S	N/S	N/S	782	MNL
Cherchi, E. & Ortuzar, JdeD. (2006)	Orthogonal	Labelled	Blocked	2	7	27	9	7 ⁴	338	3042	NL (SP-RP)
Espino, R., Román, C. & Ortúzar, JdD. (2006)	Orthogonal	Labelled	Blocked	2	5	27	9	2 ¹ ×3 ⁴	97	871	NL (SP-RP)
Greene, W.H, Hensher, D.A. & Rose, J.M. (2006)	D-efficient*	Labelled	Blocked	3-5	46	60	10	4 ⁴⁶	184	1840	MMNL
Hensher, D.A. (2006)	D-efficient	Unlabeled	Blocked	3-5	3-6	12-30	6-15	Varying	427	4593	MNL/HEV

Table 1: Review of recent stated preference studies and sample sizes in the literature¹

Reference	Design Type	Labelled/ Unlabeled	S assignment	J	к	S	Shown	design size	N	N.S.	Models Estimated
Hollander, Y. (2006)	Random	Unlabeled	N/A	2	7	N/A	9	N/A	244	2196	MNL
Loo, B.P.Y., Wong, S.C. & Hau, T.D. (2006)	Orthogonal	Unlabeled	NS	2	13	N/S	2	2 ¹ ×3 ¹³	483	815	MNL
Rizzi, L.I. & Ortuzar, JdeD. (2006)	Orthogonal	Unlabeled	Blocked	2	3	27	9	3 ³	342	3078	Binary Logit
Scott, D. & Axhausen, K.W. (2006)	Orthogonal	Labelled	Blocked	4	6	N/S	8	2 ¹² ×3 ¹²	163	1304	Probit
Scott, D. M. & Axhausen, K.W. (2006)	Orthogonal	Labelled	Blocked	4	24	72	8	2 ¹³ ×3 ¹¹	N/S	1034	MNL
Washbrook, K. Wolfgang, H. & Jaccard, M. (2006)	Orthogonal	Labelled	Blocked	3	4	32	8	3 ⁴	548	4384	MNL
Espino, R., Ortuzar, JdeD. & Roman, C. (2007)	Orthogonal	Labelled	Blocked	2	5	27	9	3 ⁵	97	871	NL (SP RP)
Fosgerau, M. (2007)	Random	Unlabeled	N/A	2	2	N/A	8	N/A	N/S	15,451	WTP space MNL
Fowkes, T. (2007)	Adaptive	Unlabeled	N/A	4	9	N/A	10	N/A	49	N/S	Weighted Regression
Hensher, D.A. & Rose, J.M. (2007)	D-efficient*	Labelled	Blocked	3-5	46	60	10	4 ⁴⁶	184	1840	MMNL
Hunt, J.D. & Abraham, J.E. (2007)	Random	Unlabeled	Random	2	5	N/A	1	Random	1128	1128	MNL
Pucket, S.M., Hensher, D.A., Rose, J.M. & Collins, A. (2007)	D- efficient¶	Unlabeled	Blocked	2+rp	7	40	4/8	5 ⁷	102	1248	MMNL
Pucket, S.M., Hensher, D.A., Rose, J.M. & Collins, A. (2007)	D- efficient¶	Unlabeled	Blocked	2+rp	7	40	4/8	5 ⁷	102	1248	MMNL
Tilahun, N.Y., Levinson, D.M. & Krizek, K.J. (2007)	Adaptive	Unlabeled	N/A	2	2	N/A	5	N/A	167	835	GLMM
Ahern, A.A. & Tapley, N. (2008)	Orthogonal	Labelled	Random	2	5	16	5/10	2 ² ×4 ³	189	1375	Rank ordered Logit
Beuthe, M. & Bouffioux, C. (2008)	Orthogonal	Unlabeled	All	1	5	25	25	5 ⁶	N/S	N/S	Ordered Logit
Hensher, D.A. (2008)	D-efficient	Unlabeled	Blocked	2+rp	5	32	16	4 ⁵	222	3552	MMNL
Hensher, D.A. (2008)	D- efficient†	Unlabeled	Blocked	2+rp	5	32	16	4^5	205	3280	MMNL
Hensher, D.A., Rose, J.M. & Black, I. (2008)	D-efficient	Unlabeled	Blocked	4	6	32	4	$2^{3} \times 4^{3}$	209 pairs	3232	MMNL
Teichert, T. Shehu, E. & von Wartburg, I. (2008)	Orthogonal	Unlabeled	Blocked	2	7	128	8	2 ⁷	5829	46,632	LCM
Train, K. & Wilson, W.W. (2008)	Random	Labelled	N/A	2	3	N/A	3	6 ³	181	543	MMNL
Tseng, Y.Y. & Verhoef, E.T. (2008)	Orthogonal	Labelled	N/S	4	15	N/S	11	$4^{13} \times 2^2$	1005	12,265	MMNL
Dagsvik, J.K. & Liu, G. (2009)	Random	Labelled	All	3	8	15	15	2 ² x3 ¹ x5 ² x7 ¹ x11 ¹ x15 ¹	100	1491	Ordered GEV
Hensher, D.A., Rose, J.M. Ortúzar, JdeD &.Rizzi, L.I. (2009)	D-efficient	Unlabeled	Blocked	2	12	60	10	3'x5°x6² x10 ¹ x20 ¹ x30 ¹	213	2130	MMNL
Hess, S. & Rose, J.M. (2009)	D- efficient†	Unlabeled	Blocked	2+rp	5	32	16	4 ⁵	205	3280	MMNL
Lee, J. & Cho, Y. (2009)	Orthogonal	Unlabeled	Blocked	Ranked	5	216	20	2 ¹ ×3 ³ ×4 ¹	492	9840	Linear Regression
McDonnell, S. Ferriera, S. & Convery, F. (2009)	Orthogonal	Unlabeled	Random	2+rp	7	32	18	$2^{2} \times 3^{2} \times 5^{3}$	310	5442	MMNL
Puckett, S.M. & Hensher, D.A. (2009)	D-efficient	Unlabeled	Blocked	2+rp	7	40	4	57	108	432	MMNL
Sener, I.N., Eluru, N. & Bhat, C.R. (2009)	Orthogonal	Unlabeled	Blocked	3	6	N/S	4	$2^{1} \times 3^{8} \times 5^{2}$	1621	6484	MMNL
Rouwendal, J. de Blaeij, A., Rietveld, P. & Verhoef, E. (in press)	Orthogonal	Unlabeled	Blocked	2	3	N/S	N/S	N/S	1055	N/S	MMNL

*, ¶, † indicate that the same data set has been used in these studies, N/S = not stated, N/A = not applicable

2.2 A Review of the Empirical Literature Exploring the Impact of Experimental Designs upon Modelling Outcomes

A small number of studies have recently begun to examine empirically the differences between data sets collected using efficient and orthogonal designs. Louviere et al. (2008) in a single study examined 44 different experimental designs varying systematically the various design dimensions as well as the level of statistical efficiency of the designs. It was found that the more statistically efficient the design, the greater the error variance resident within the data. This relationship was found to exist independent of the overall dimensions of the design. Hess et al. (2008) compared the results of three different experimental designs, including an orthogonal design with randomized choice set assignment, an orthogonal design with an orthogonal blocking column and an efficient design. They found that the efficient design performed only marginally better than the orthogonal design with blocking, but that the design with random assignment of choice tasks to respondents performed significantly worse than both the efficient design and the orthogonal design with blocking. As such, they concluded that the blocking of the experiment was far more important than the underlying experimental design itself.

Whilst contributing to our understanding of the impacts differing experimental design types might have upon modeling outcomes, both of these studies fail to examine whether the two main theoretical advantages commonly associated with efficient designs actually translate into practice. That is, efficient designs have been promoted within the literature as producing more reliable parameter estimates given smaller sample sizes than other forms of designs, such as orthogonal designs. It is this claim that we seek to examine here.

3. EMPIRICAL STUDY: AIR TRAVEL CHOICE

3.1 Empirical Context

The study presented in this paper deals specifically with air travel choice. Typically, when travelling to (larger) cities, it is possible to fly to a main airport close to the city itself. However, flying into main airports is becoming more expensive, both for the individual traveler and the airlines themselves. For this reason, there has been a growth in the use of regional airports that are located in the proximity of major cities. Such airports typically are cheaper to fly to (due to lower taxes) but require greater travel distanced to either access or egress. In the current study, we examine the trade-offs mainly between ticket price and travel time. One aim of the study is to capture the behaviour of respondents by presenting choices that closely resemble real-life choices. For this reason, we selected a particular trip, namely a holiday trip from Amsterdam to Barcelona, a popular holiday destination with a significant frequency of trips. The origin airport for the study is Schiphol Airport in Amsterdam, while the destination airport is either the main Barcelona Airport or the regional Girona Airport. The latter airport is much further away from Barcelona city, being 100 kilometers distant from Barcelona compared to 13 kilometers for Barcelona Airport (see Figure 1).

Currently, several charter airlines offer cheap tickets to Girona Airport, such that the traveller has the option to pay less for the air component of travel, but requires a longer travel time to Barcelona city by bus. Travellers choose a ticket among others based on ticket price, departure time, number of stops or transfers, and airline company (one can have a preference for a certain airline based on for example, experienced comfort or air miles membership card). In the current study, we also take destination airport and travel time and cost to the city centre into account.



Figure 1: Main airport Barcelona Airport and regional airport Girona Airport

3.2 The Discrete Choice Experiment

Respondents were asked to complete an online DCE involving the choice of air ticket from amongst a set of five possible tickets. In the experiment, each of the five alternative tickets were described by the name of the airline offering the ticket, the ticket price, the departure time, the number of transfers required and the destination airport. To capture egress times and costs, the various tickets also displayed the time required to travel from the destination airport to the city of Barcelona as well as the price. Given that the flight time from Amsterdam to Barcelona airport and Girona airport are roughly the same, this was fixed at two hours and not varied as part of the experiment.

Each respondent was exposed to six choice tasks in which the attributes shown were varied according to some underlying experimental design (see Section 3.3). The levels that each ticket could take as part of the experiment are shown in Table 2. Note that depending on the destination airport, different egress time and cost levels were selected. It should be noted that all attribute levels are consistent with a single trip, and that for example for a round trip double the price is to be considered. Figure 2 shows an example choice task given to respondents (two alternatives are cut off in the screenshot).

Within each choice task, respondents were able to sort the presented alternatives by airline, price, travel time, etc., and also select which attributes to show or omit. There is also an option to show the total travel time (flight time plus transfer time plus egress time) and total price (flight price plus egress price). This flexibility was included so as to enhance the feeling of dealing with an actual online travel agent rather than a traditional DCE. Of course, before the choice situations are presented, the scenario is sketched (a holiday trip from Amsterdam to Barcelona), and in the end also questions about the respondent (gender, income, etc.) are asked.

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

Table 2: The attribute levels used as part of the DCE

* If destination is Barcelona, other values are for destination Girona.

Attribute		Show	Sort by	Informatio	n				
Flight									
Airline		V	0						
Departure Time		V	\bigcirc						
Arrival Time		V	0	Depature tin	ne + infligh	t time + transfer	time		
Number of Stops / Trans	sfer Time	V	O	Time spent	waiting at t	he stop(s)			
Travel Time		V	O	Travel time r (including in	needed fro flight time	m Amsterdam to and transfer time) Barcelona airp e)	ort or to	Girona airport
Price			۲	Ticket price	(flight only)	including fees a	and taxes		
Ground transportation									
Travel Time		V	0	Additional p	rice for trav	eling from the ar	rival airport to th	ne city ce	ntre
Price		V	\odot	Additional tir	me needeo	to travel from th	e arrival airport	to the cit	ty centre
Total Travel Time		V	0	Total travel time (including inflight time and ground transportation travel time) from Amsterdam airport to Barcelona City					
Total Price			O	Total price (i Amsterdam	including fl to Barcelo	ight price and gr na City	ound transporta	ation pric	e) from
Iberia <i>IBERIA</i>						Total	3 hr 40 min	€ 65	
Flight	From		To Circono Airport	Departure e:oo	Arrival	Stops	Travel Time	Price	Choose this ticket
Ground Transportation	Girona Airport		Barcelona City	0.00	0.00	0	2 m 00 min 1 hr 40 min	€ 15	
			,						
Transavia (transavia.com						Total	5 hr 00 min	€65	
	From		То	Departure	Arrival	Stops	Travel Time	Price	Choose this ticket
Flight	Amsterdam		Girona Airport	12:00	16:00	1 - (2 hr 00)	4 hr 00 min	€ 50	Choose and leker
Ground Transportation	Girona Airport		Barcelona City				1 hr 00 min	€15	
March 199						T -1-1	4 1- 00		
Vueling voeing com			_			iotai	4 nr 20 min	€84	
Clink4	From		To Oissue Aissuet	Departure	Arrival	Stops	Travel Time	Price	Choose this ticket
Flight Ground Transportation	Amsterdam Girona Airport		Girona Airport Barcelona City	0:00	9:00	n - (n nr 00)	3 nr 00 min 1 hr 20 min	€/5 €0	
oround transportation	Girona Airport		Darcelona City				110/201000	69	

Figure 2: Example choice task

3.3 The Experimental Design Procedure

Three different experimental designs have been generated; one orthogonal design and two D-efficient designs. The smallest (simultaneous) orthogonal design that could be located with all attribute levels uncorrelated had 108 choice situations. Given that it was determined that presenting each respondent with 108 choice situations was not feasible, an orthogonal blocking column was generated such that 18 blocks were created, each containing six different choice situations. Alongside the orthogonal design, a D-efficient design was constructed with 108 choice situations which was also blocked into 18 blocks containing six choice tasks each. Finally, a second D-efficient design was constructed with 18 choice situations which was subsequently blocked into three blocks. In the final survey, each respondent faced six choice situations from one of these three designs. The block allocation to the respondents was implemented in such a way that an even distribution over the different designs was established. The actual experimental designs used are available from the authors upon request.

In order to construct MNL efficient designs, the analyst is required to assume prior parameter estimates in a Bayesian-like fashion. Researchers need not assume

precise prior parameter values but rather, may assume prior parameter distributions that are expected to contain within their range, the true population parameter. In taking this approach, the resulting Bayesian efficient design is then optimised over a range of possible parameter values, without the analyst having to know the precise population value in advance (see e.g., Sandor and Wedel 2001 and Kessels et al. 2006). For the current study, the two D-efficient designs were generated using parameter priors obtained from a small pilot study consisting of 36 respondents yielding a total of 216 choice observations. The pilot experiment used the aforementioned orthogonal design such that each block was replicated exactly twice. The parameter estimates of the pilot study have been estimated according to a MNL model where the attribute 'Airline' has been dummy-coded with base 'EasyJet', and departure time has been dummy-coded with base '6pm'.

Table 3 shows the model results for the pilot study. The prior parameters used for design generation were drawn from Bayesian normal distributions $\beta \sim N(\mu, \sigma^2)$ where the mean μ was assumed to be the estimation value of the parameter from the pilot study and the standard deviation σ the standard error of the parameter. For non-significant parameter estimates, the prior parameters were assumed to be zero with the corresponding standard error as value for the standard deviation (e.g. the prior for egress price is assumed to be normally distributed with mean 0 and standard deviation 0.03).

Attribute	Parameter	<i>t</i> -value	Prior (mean)	Prior (stdev)
Air France	-0.516	-1.596	0	0.323
KLM	0.194	0.661	0	0.293
Iberia	-0.218	-0.729	0	0.299
Vueling	0.125	0.426	0	0.293
Transavia	-0.547	-1.765	0	0.310
Ticket Price	-0.040	-8.444	-0.040	0.005
Departure Time (6am)	0.215	0.974	0	0.221
Departure Time (12pm)	0.614	2.883	0.614	0.213
Transfer Time	-0.006	-3.298	-0.006	0.002
Egress Price	-0.027	-0.904	0	0.030
Egress Time	-0.018	-3.265	-0.018	0.006

Table 3: Pilot study parameter estimates and efficient design priors

 ρ^2 =0.190, LL=-266

3.4 Main Field Phase Sample

Respondents were sampled by a Dutch market research company, TeamVier, who sent out emails to people in their panel database. In total 618 respondents completed the survey, such that the total number of observations (chosen alternatives) is 3,708. Of this sample, 206 (1,236 observations) respondents completed choice tasks from the orthogonal design, 208 (1,248 observations) from the efficient design with 108 choice tasks, and 204 (1,224 observations) from the smaller efficient design.

The age groups 26-35, 36-45, and 60+ are more or less equally represented in the sample, while age group 45-59 is the largest group and 18-25 is the smallest group. Slightly more males than females are in the sample. Most people have a full time job,

the second largest group has a part time job, another large group has no job, and a small group stated to be a student. A quite uniform distribution of income is observed, from ≤ 0 to $\leq 70,000$ per year.

4. EMPIRICAL STUDY: AIR TRAVEL CHOICE

4.1 Overall Model Results

Before examining the impact of the different experimental designs on sample size, several models were estimated both on the full pooled data from all three experimental designs (i.e., using all 3,708 choice observations), as well as on the separate design specific data sets. Table 4 presents the MNL model results from these model runs. As well as estimating MNL models, the 'nested logit trick' was used to test for scale differences between each of the data sets. The result of this model are not reported however as the model collapsed back to the MNL model, suggesting that the different designs do not induce different degrees of error variance. It should be noted that this finding does not contradict that of Louviere et al. (2008) however as in that paper, they explored not only different designs, but designs with differing design dimensions. In the current study, the design dimensions were kept constant with only the experimental design varying.

Table 4 lists the parameter estimates on data from each design separately and pooled. Looking at the airline parameters (relative to 'EasyJet') of the pooled data, the order of preference for airlines appears to be (from most preferred to least preferred): Air France, EasyJet, Transavia, Vueling, KLM, and Iberia. As expected, both price parameters are negative, and so are the transfer time and egress time parameters. It is interesting to look at the parameter ratios, indicating WTP for less transfer time or less egress time. The WTP for one hour less transfer time is approximately \in 32.14 (based on the flight price) compared to a WTP for one hour less egress time is \in 19.29 (based on the flight price) or \in 15.43 (based on the egress price). Therefore, roughly speaking one is willing to pay about \in 10.29 (single trip) for flying to Barcelona instead of Girona airport, which is about 40 minutes further away.

Attribute	parameter	t-ratio	lower 95%	upper 95%
Air France	0.377	6.019	0.254	0.500
KLM	-0.232	-3.456	-0.364	-0.101
Iberia	-0.535	-7.629	-0.672	-0.398
Vueling	-0.008	-0.120	-0.132	0.117
Transavia	-0.137	-2.046	-0.268	-0.006
Ticket Price	-0.028	-25.115	-0.030	-0.026
Departure Time (6am)	0.283	6.221	0.194	0.372
Departure Time (12pm)	0.460	9.860	0.368	0.551
Transfer Time	-0.015	-32.228	-0.016	-0.014
Egress Price	-0.035	-5.956	-0.047	-0.024
Egress Time	-0.009	-8.459	-0.011	-0.007
Model fits				
LL(ASC only model)	-5879.526			
LL(β)	-4774.862			
ρ^2	0.188			
Respondents	618			
Observations	3708			

Table 4: MNL model results (pooled data)

Note that the parameters in Table 4 are different from the parameters used as priors for the generation of the efficient designs in Table 3. Efficient designs are efficient under the assumption that the priors are correct, hence it is likely that some efficiency is lost in the data, although the data from a Bayesian efficient design remains relatively efficient over a range of parameter prior values.

In terms of testing the hypothesis put forward by Hess et al. (2008) with relation to blocking representing the most important criteria of experimental design, given that all designs in the current study were blocked, we are unable to test at the level of estimating models on the full design specific data sets, the impact of blocking upon the design specific results. We are able to test this impact however in Section 4.4 where we use bootstrapping at various sample sizes, and compare the results after i) maintaining equal representation of respondents over the blocks, and ii) bootstrapping without reference to the blocks that respondents were assigned.

4.2 Design Specific Model Results

Table 5 presents the MNL model results for each of the designs. The statistically significant parameter estimates have the same sign across the three design types and are more or less of the same magnitude. However, there are some notable differences. For example, the parameter estimates for transfer time and ticket price, the two parameters with the highest t-ratios, are guite different between the orthogonal and the two efficient designs, and their ratio even more so. Although asymptotically any design type should lead to the same parameter estimates, in limited samples the parameter estimates can clearly be different. While one would expect the parameter estimates to be unbiased (in large samples), regardless of the design type used, the design could have an influence. When a design would consist of many dominant alternatives, the parameters are likely to be larger (in an absolute sense), as the error variance will be smaller for such choice situations. Hence, in a stated choice experiment the parameters may become biased when many dominant alternatives are included in the experimental design. As the orthogonal design is generated without any assumptions on parameter priors, one cannot avoid dominant alternatives here (unless one removes them by manual inspection), whereas efficient designs mostly rule out choice situations with dominant alternatives (as such a choice situation would be very inefficient). Inspection of the orthogonal design indeed shows that it contains several choice situations with clearly dominant alternatives, whereas the efficient designs contain very few (efficient design with S=108) or none (efficient design with S=18). In Bliemer and Rose (2009a) evidence is found that bad priors may lead to dominant alternatives in the design, biasing the parameter estimates, at least in small sample sizes.

4.3 Comparison of Design Outcomes against Model Results

In generating the design, we assumed some parameter priors obtained from a small pilot study, see Table 3. The efficient designs were generated under these assumptions. The estimated parameter values are presented in Table 4 (and Table 5 for each design separately). There are some obvious differences between the assumed means of the priors and the estimated parameters, such that the efficient designs likely loose some efficiency. However, we have tried to minimize this loss in efficiency by adopting Bayesian efficient design principles in which we included the uncertainty about these priors (given by the standard deviations in Table 3), as explained in Section 3.3.

Table 5: MNI	_ model res	sults for	different	designs
--------------	-------------	-----------	-----------	---------

	Desigr	Design 1: Orthogonal, S = 10				Design 2: Efficient, S = 108				Design 3: Efficient, S = 18			
			lower	upper			lower	upper			lower	upper	
Attribute	par.	t-ratio	95%	95%	par.	t-ratio	95%	95%	par.	t-ratio	95%	95%	
Air France	0.422	3.423	0.180	0.664	0.319	2.994	0.110	0.527	0.457	4.401	0.253	0.660	
KLM	-0.231	-1.883	-0.471	0.010	-0.402	-3.265	-0.643	-0.161	-0.112	-1.017	-0.328	0.104	
Iberia	-0.488	-3.756	-0.742	-0.233	-0.650	-5.445	-0.884	-0.416	-0.417	-3.523	-0.649	-0.185	
Vueling	0.285	2.345	0.047	0.524	-0.141	-1.324	-0.349	0.068	-0.131	-1.173	-0.351	0.088	
Transavia	0.024	0.190	-0.227	0.276	-0.301	-2.635	-0.524	-0.077	-0.052	-0.469	-0.270	0.166	
Ticket Price	-0.035	-17.780	-0.038	-0.031	-0.023	-10.808	-0.027	-0.019	-0.021	-9.649	-0.026	-0.017	
Departure Time (6am)	0.300	3.311	0.123	0.478	0.410	5.387	0.261	0.560	0.171	2.285	0.024	0.318	
Departure Time (12pm)	0.627	7.181	0.456	0.798	0.348	4.181	0.185	0.511	0.303	3.786	0.146	0.460	
Transfer Time	-0.011	-13.999	-0.012	-0.009	-0.015	-19.410	-0.017	-0.014	-0.018	-22.350	-0.020	-0.017	
Egress Price	-0.041	-3.311	-0.065	-0.017	-0.038	-3.894	-0.057	-0.019	-0.032	-3.330	-0.051	-0.013	
Egress Time	-0.012	-5.205	-0.016	-0.007	-0.006	-3.102	-0.010	-0.002	-0.006	-3.342	-0.01	-0.003	
Model fits													
LL(ASC only model)		-1953	577			-1940	.001			-1983	.633		
LL(β)		-1557	.904			-1607	.444			-1550	.317		
ρ ²		0.2	03			0.1	71			0.2	18		
Respondents		20	6			20	4			20	8		
Observations		123	36			124	48			122	24		

In the design generation process, the predicted AVC matrices are used to assess their efficiency. In model estimation the variance-covariance matrices are an output. These variance-covariance matrices determine the standard errors and the *t*-ratios. In this section we would like to compare the predicted *t*-ratios in the design generation process (before the survey) with the *t*-ratios computed at model estimation (after the survey). The model *t*-ratios are listed in Table 5. The *t*-ratios for each design can be computed by taking the square roots of the diagonals of the in the design process predicted asymptotic variance-covariance matrices for the given sample sizes (206, 204, and 208 respondents for the three designs, respectively). In order to be able to make the comparison, the parameter estimates from the pooled data set are used priors.

Figure 3 depicts the comparison between the predicted and observed (in estimation) *t*-ratios for all parameters in each of the three datasets (hence, 11x3 = 33 points in the figure). Overall, the prediction is quite good, where predicted low *t*-ratios are observed as low, and high predicted *t*-ratios are observed high. A very interesting observation is, that the observed *t*-ratios seem to be slightly smaller than the predicted *t*-ratios. Hence, in this case study the predicted *t*-ratios can be seen as a lower bound, where the bound is general quite tight, although some outliers exist. Therefore, we conclude that the theoretical design principles are indeed very useful and quite accurate in making predictions about standard errors and *t*-ratios without conducting any survey. Hence, it is possible to identify parameters in advance that will likely not be statistically significant in estimation. Also, sample sizes can be derived as discussed in Bliemer and Rose (2009b). Note that the predicted *t*-ratios will deviate more if the assumed priors deviate more from the actual parameter estimates. Therefore, it is wise to invest in a good pilot study in order to obtain a good prediction of the *t*-ratios.



Figure 3: Predicted (before survey) and observed (after survey) t-ratios (red = orthogonal design, blue = efficient design 1 (S=108), purple = efficient design 2 (S=18)

4.4 Comparison of the Efficiencies of the Design Specific Data Sets

The theoretical justification for using efficient as opposed to orthogonal designs is that the former are expected to produce smaller standard errors at a given sample size, or conversely require smaller sample sizes to produce larger *t*-ratios. To date, and to the best of our knowledge, this has only been shown to occur in simulated data sets. In this section, we examine whether the theoretical advantages of generating efficient designs actually translate to real life empirical data sets. We will analyze our three data sets that resulted from the three underlying experimental designs by (i) looking at the aggregate results using all respondents, and (ii) bootstrapping on each data set in order to randomly select respondents in the dataset for different sample sizes.

Using the aggregate results, we can compute the D-error for each of the data sets, which is the determinant of the variance-covariance matrix, normalized to the number of parameters. The D-errors for each of the data sets are presented in Table 6. In order to compare these D-errors, we have further normalized each D-error to a single respondent. The lower this D-error, the more efficient the data set. As the table shows, the two efficient designs generated almost equally efficient data sets. Hence, a large design with 108 choice situations need not outperform a small design with only 18 choice situations. Since the number of choice situations in the orthogonal design could not be decreased (as we were not able to find any such orthogonal design), this is a clear advantage of the efficient designs in which the number of choice situations can be kept small. Also from Table 6 we can conclude that the two efficient designs indeed outperform the orthogonal design in terms of efficiency, which should translate itself to smaller standard errors, will is investigated next.

	Orthogonal design	D-efficient design 1	D-efficient design 2
	(S=108)	(S=108)	(S=18)
D-error	0.1253	0.1035	0.1017

Table 6: Comparison of D-errors

We use bootstrapping to examine the performance of the three designs over different sample sizes. Simultaneously, we also test the impact that blocking has upon design performance by performing two sets of bootstrapping on each. Firstly, we bootstrap in such a way that each block in each design is replicated an equal number of times such that the properties of the design will be maintained through to model estimation. Next, we perform bootstrapping in such a way that respondents are randomly selected, independent of the design block to which they were exposed. In this way, each bootstrap iteration will likely sample respondents unequally from each block, thus impacting upon the design properties in terms of how they are expected to translate through to model estimation. In each case, we perform 250 bootstrap iterations using the same utility specification that was assumed previously.

First, we examine the impact upon the standard errors allowing for equal representation within each of the blocks. Across all sample sizes, at the 95 percent confidence level, seven of the 11 standard errors for the orthogonal design were found to be statistically different when comparing the results of maintaining equal block representation against unequal representation, only three were statistically different for the first efficient design, whilst all 11 were statistically different for the second efficient design. This suggests that blocking may have a statistically significant impact upon the statistical efficiency of the parameter estimates, although this impact appears to be small in magnitude. Therefore, in the analysis following we will just consider the results in which each block is represented an equal number of times, noting that the results for unequal representations of the blocks are similar.

Figure 4 graphs the average standard error of each parameter over a range of sample sizes based on the bootstrap simulations. The graphs show that the two efficient designs produce very similar standard errors, which are in smaller than the standard errors produced by the orthogonal design. Only in two situations does the orthogonal design produce smaller standard errors, namely for the ticket price parameter and (to a lesser extent) the transfer time parameter. This is no coincidence. In the graphs also the levels of standard error required to obtain a statistically significant parameter estimate (t = 1.96) are plotted by horizontal solid black lines, based on the pooled parameter estimates. The ticket price and transfer time parameters can easily be estimated with a very small sample size. Efficient designs are optimized such that they trade off all standard errors simultaneous, such that all parameters can be estimated with relatively low standard errors. It is clear that the ticket price and transfer time parameters already have low standard errors, hence the efficient designs mainly aim to minimize the standard errors of the other parameters, accepting slightly higher standard errors for ticket price and transfer time.

The graphs for the two efficient designs are very similar, again finding that an efficient design with only a small number of choice situations is just as good as efficient designs with a large number of choice situations. Not only is it just as good, in larger sample sizes it even (slightly) outperforms the larger efficient design, while in smaller sample sizes the larger design seems to yield smaller standard errors. This is an interesting observation, which may be explained by the fact that for small sample sizes (with only few respondents) the variation in the choice situations may play a dominant role, while for large sample sizes the (few) best choice tasks, present in the smaller design, yield best results.



Figure 4: Standard errors for different designs per sample size

The improvement in standard errors when using an efficient design may seem small, but standard errors that are a factor p higher require a factor p^2 higher sample size. For example, if a standard error is 10 percent higher (factor 1.1), then a 21 percent higher sample size is required. Table 7 presents the sample sizes (required for statistical significant parameter estimates at the five percent level) for the nine statistically significant parameter estimates corresponding to the graphs in Figure 4. The percentages in Table 7 are deviations from the sample sizes required for the best design, the second efficient design, which performs best for each parameter. The first efficient design performs almost equally well. The orthogonal design requires for all but one parameter significantly higher sample sizes, ranging from 16

percent to 61 percent more respondents for the same level of statistical significance. The parameter that requires less respondents, the ticket price parameter, is only a decrease from 5 to 4 respondents and therefore not that relevant.

	Orthogona (S=1)	Orthogonal design (S=108) sample size increase*		design 1 08)	D-efficient design 2 (S=18)
Attribute	sample size	increase*	sample size	increase*	sample size
Air France	85	+39%	63	+3%	61
KLM	221	+23%	221	+23%	179
Iberia	47	+21%	39		39
Ticket Price	4	-20%	5		5
Departure Time (6am)	82	+46%	57	+2%	56
Departure Time (12pm)	29	+16%	26	+4%	25
Transfer Time	3		3		3
Egress Price	98	+61%	61		61
Egress Time	47	+47%	37	+16%	32

Table 7: Compansion of sample sizes	Table	7:	Com	parison	of	sam	ple	sizes
-------------------------------------	-------	----	-----	---------	----	-----	-----	-------

* extra sample size required, compared to D-efficient design 2.

5. DISCUSSION AND CONCLUSIONS

In this paper, we have provided a systematic and detailed study comparing empirically various claims made by researchers as to the benefits of using efficient experimental designs for studies involving DCEs. In it, we have provided an overview of the various types of experimental designs that have been used in practice for the past decade based on a literature review of several tier one transportation journals. Based on this literature review, it is clear that the use of efficient designs is growing within the literature, although their use still remains in the minority.

In this study we showed that the (asymptotic) variance-covariance matrices computed in the design generation process (before the survey) closely corresponds to the variance-covariance matrix in estimation (after the survey), such that for example *t*-ratios can be predicted with a quite high level of accuracy for each parameter before the experiment is taken to field. However, the prediction quality depends on the accuracy of the parameter priors needed to generate the efficient designs, making pilot studies even more essential.

An important finding in this research is that we could indeed empirically produce lower standard errors with efficient designs compared to an orthogonal design as suggest by the theory. These lower standard errors translate into significant decreases in sample size for achieving the same level of statistical significance in estimation. Also, we found that efficient designs with a large number of choice situations need not be more efficient per choice situation than an efficient design with a small number of choice situation, even the contrary may be true.

When we look at the parameter estimates themselves, different designs yield somewhat different parameter values and the question rises of the design can bias the parameter estimates. It is not entirely clear why this is the case, but we conjectured that it may lie in the fact that dominant alternatives in a stated choice study bias the parameter estimates. While the orthogonal design contains the most choice situations with dominant alternatives, they may be the most biased. An alternative possible cause might lie in the way that orthogonal designs are constructed. These designs are constructed so that each pairwise attribute combination appears an equal number of times (or as close to possible) over the design. Thus for example, the 6am, 12pm and 6pm departure time levels each appear 12 times with the 0, 60 and 120 minute transfer time attribute levels. This relationship need not, indeed does not, hold for the efficient designs. As such, it is

possible that in a stated choice experiment, the parameters may become biased when using efficient experimental designs in practice. At present, this remains conjecture, and urgent research is required to confirm or rule out this as a possibility.

The current paper suffers from a number of limitations, the most predominate of which is our use of MNL models. As has been well know for some time, the MNL model fails to account for the correlation of preferences within individuals across choice tasks. To account for this phenomenon, more advanced models such as MMNL or error components models are required. Whilst we acknowledge this limitation here, we note that previous research has found that designs generated specifically for MNL models, tend to perform reasonably well when more advanced models are used (see e.g., Bliemer and Rose 2008 or Rose et al. 2009). We note however that this research also tends to rely on model simulations and hence empirical work is required to confirm whether this finding also translates into practice.

Overall, we were unable to substantiate the findings of Louviere et al. (2008) in terms of efficient designs tend to induce greater error variance, although we have also conjectured here that that might indeed be the case. We hypothesis that this might be due to orthogonal designs producing far more dominated alternatives than efficient designs, although again, this requires further research to confirm. Our findings do however tend to support those of Hess et al. (2008) in that blocking of the design appears to be of some importance. Our study differs to that of Hess et al. (2008) in which we do not randomly assign choice tasks to respondents but rather maintain respondents within blocks. We further distinguish our research to that of Hess et al. (2008) by comparing model results of maintaining between block representativeness to a random assignment of respondents independent of blocks, nevertheless, we reach similar conclusions in terms of the importance of blocking and maintaining equal sampling over blocks in DCEs.

In conclusion, our research empirically supports some of the claims made by those researching in the area of efficient designs. Despite the above acknowledged shortcomings, we have found that these designs do indeed appear to produce more reliable estimates. Whilst we recommend more studies similar to the one described here, encompassing greater numbers of types of designs be carried out, we find no reason at this point to reject the use of efficient designs, and indeed we recommend their use in studies where smaller sample sizes are likely to be a reality.

Notes

- 1. Journals examined are limited to (in alphabetical order) Journal of Transport Economics and Policy, Transportation, Transportation Research Part A, Transportation Research Part B, and Transportation Science.
- 2. Some studies used a subset of observations from the total sample population. Numbers reported are those used in the estimation of models within the article.
- 3. In many papers reviewed, the authors were unable to determine the dimensions of the experimental design or information about the sample size and number of observations captured. Where we were unable to determine these facts, we have marked the appropriate cell in Table 1 as 'not stated'. The conclusion that can be drawn from our literature review is that much improvement can be made in terms of what is reported within the SC literature.

References

Bliemer, M.C. and Rose, J.M. (2005) Efficiency and Sample Size Requirements for Stated Choice Studies, Working paper: ITLS-WP-05-08.

Bliemer, M.C.J. and Rose, J.M. (2006) Designing Stated Choice Experiments: Stateof-the-art, paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, Japan.

Bliemer, M.C. and Rose, J.M. (2009a) Efficiency and Sample Size Requirements For Stated Choice Experiments, *Transportation Research Board Annual Meeting*, Washington DC January.

Bliemer, M.C.J. and Rose, J.M. (2009b) Serial Choice Conjoint Analysis for Estimating Discrete Choice Models, *International Choice Modelling Conference*, Leeds, UK.

Bunch, D. S., Louviere, J. J. and Anderson, D. A. (1996) A comparison of experimental design strategies for multinomial logit models: the case of generic attributes. Working paper, Graduate School of Management, University of California at Davis.

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

Ferrini, S. and Scarpa, R. (2007) Designs with a-priori information for nonmarket valuation with choice-experiments: a Monte Carlo study, Submitted to *Journal of Environmental Economics and Management*, 53, 342-363.

Fowkes, A. S. 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, 27–44.

Hess, S., Daly, A. Rohr, C. and Hyman, G. (2007) On the development of time period and mode choice models for use in large scale modelling forecasting systems, *Transportation Research Part A*, 41(9), 802-826.

Hess, S., Smith, C., Falzarano, S. & Stubits, J. (2008) Measuring the effects of different experimental designs and survey administration methods using an Atlanta Managed Lanes Stated Preference survey, *Transportation Research Record*, 2049, 144-152.

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

Kanninen, B.J. (2002) Optimal Design for Multinomial Choice Experiments, *Journal of Marketing Research*, 39 (May), 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), 409-419.

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

Louviere, J.J., Islam, T., Wasi, N., Street, D. and Burgess, L. (2008) Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price? Journal of Consumer Research, 35(2), 360-375.

Louviere, J.J. and Woodworth, G. (1983) Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data, *J Marketing Research*, 20, 350-367.

Rose, J.M. and Bliemer, M.C.J. (*in press*) Constructing Efficient Stated Choice Experimental Designs, *Transport Reviews*.

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, *International Choice Modelling Conference*, March 30-April 1, Yorkshire U.K.

Rose, J. M. and Bleimer, M. C. J. (2008) Stated preference experimental design strategies, in: D. A. Hensher and K. J. Button (Eds) *Handbook of Transport Modelling*, Ch 8, 151–180 (Oxford: Elsevier).

Sándor, Z. and Wedel, M. (2001) Designing Conjoint Choice Experiments Using Managers' Prior Beliefs, *Journal of Marketing Research*, 38 (November), 430-444.

Sándor, Z. and Wedel, M. (2002) Profile Construction in Experimental Choice Designs for Mixed Logit Models, *Marketing Science*, 21(4), 455-475.

Sándor, Z. and Wedel, M. (2005) Heterogeneous Conjoint Choice Designs, *Journal of Marketing Research*, 42, 210-218.

Scarpa, R. and Rose, J.M. (2008) Designs efficiency for non-market valuation with choice modelling: how to measure it, what to report and why, submitted to *Australian Journal of Agricultural and Resource Economics*, 52(3), 253-282.

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, 185–199.

Toner, J. P., Clark, S. D., Grant-Muller, S. M. and Fowkes, A. S. (1999) Anything you can do, we can do better: a provocative introduction to a new approach to stated preference design, *WCTR Proceedings*, Antwerp, 3, 107–120.

Train, Ken (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.