AN EVIDENCE BASED ANALYSIS OF ORTHOGONAL AND EFFICIENT EXPERIMENTAL DESIGNS FOR TRANSPORTATION STATED PREFERENCE SURVEYS

G HAYES

Department of Civil Engineering, Centre for Transport Development,
University of Pretoria; Tel: 063 795 4887; Email: u16401868@tuks.co.za

ABSTRACT

Discrete choice models have had wide application for mode choice simulation as part of the multi-modal four-step transport demand modelling process in South Africa. For all models since 2000, urban commuter modal trip preference data for the mode choice sub-model has been collected using stated preference (SP) surveys. The SP experimental designs have been fractional factorial types (FFD). The SP data and the subsequent mode choice models were directly dependent on the soundness of the underlying experimental designs. This paper presents a review of the experimental designs of four South Africa metropolitan SP survey case studies to derive empirical evidence of the efficiency of these designs as measured by the D-error. The analysis showed there were several inadequacies in the experimental designs. Furthermore, the multinomial logit (MNL) models estimated from the SP data for each metro highlighted the shortcomings of the experimental designs. The paper illustrates that there is enough insight into prior estimates of the attribute parameters for use in efficient experimental designs and that these designs provide improved D-error measures over the equivalent orthogonal fractional factorial designs. The paper recommends that efficient designs be used in the future for all SP transport applications.

1. INTRODUCTION

This paper presents the analysis of the SP experimental designs of four South African metropolitan case studies for the development of discrete choice models for the estimation of the diversion of car drivers to the Bus Rapid Transit (BRT) mode. The designs and surveys were completed between 2011 and 2015. The case studies are for four large metropolitan areas in South Africa that were among the first to plan and operate BRT systems. This system planning process required the estimation of the BRT passenger demand and fare revenue for the system designs, operational plans and financial modelling activities. The diversion of passenger trips from the taxi, rail, bus and car modes to BRT was anticipated, and the car – BRT experimental designs and choice models are focused on in this paper. MNL models that were derived from the SP data collected subsequent to the finalization of the experimental designs are presented and discussed. Based on the attribute parameters estimated from one of these MNL models, an efficient design is presented. The paper demonstrates that this design provides better efficiency as measured by the D-error than the equivalent orthogonal fractional factorial designs and recommends that efficient designs be used in South Africa in the future.

This paper is organised as follows. Section 2 begins by describing the general principles of experimental designs, focusing initially on orthogonal fractional factorial designs. In Section 3 orthogonality is discussed as a design requirement and evidence is presented that orthogonality is rarely achieved in execution. In Section 4 the theory and application of efficient experimental designs in the transport mode choice context is discussed. The case study analysis of the four metropolitan fractional factorial experimental designs is presented in Section 5. MNL models derived from the metropolitan SP data are estimated and discussed in Section 6. In Section 7 an efficient design for a car versus BRT SP survey is presented and discussed. Section 8 draws conclusions from the analysis and provides recommendations.

2. BACKGROUND

An experiment is an investigation that establishes a particular set of circumstances under a specified protocol to observe and evaluate the implications of the resulting observations (Keuol & Burger, 2018). In SP studies this translates into the need to determine the influence of the utility attributes upon the choices that are observed to be made by respondents undertaking the experiment. To ensure the statistical integrity of the model to be estimated from the observed experimental data it is necessary for the analyst to control and quantify the effect of variations made to the independent variables (utility attributes) on the dependent variable (choice). The manipulation of the levels of the independent variables have conventionally been done for linear models in accordance with orthogonal design principles that minimise attribute level correlation. These designs are also popular because they are easy to construct and they have historical impetus (Rose & Bliemer, 2009). High levels of attribute parameter correlation can manifest by either the incorrect parameter signs (multicollinearity in linear models) and/or high parameter standard errors, rendering the models statistically inadequate. While orthogonal designs are still commonly employed for non-linear models by many practioners, they are not the most suitable (Bliemer & Rose, 2010), and require large sample sizes (Puckett & Rose, 2010).

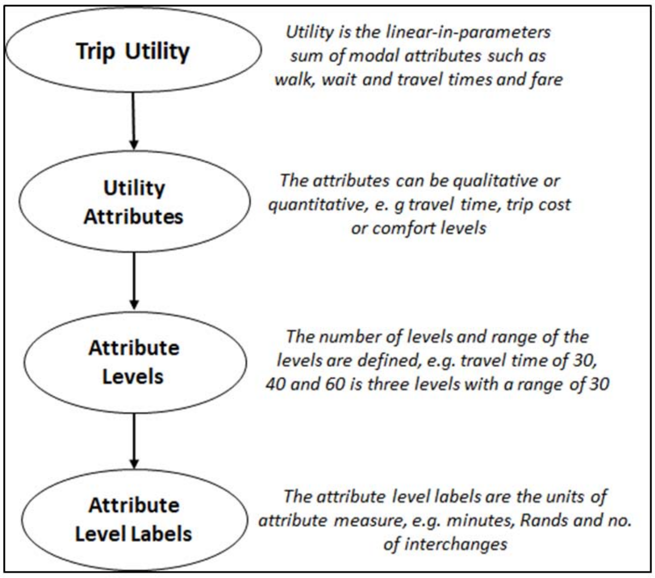

Figure 1: Hierarchical Structure of Utility Attributes, Levels and Labels

In the context of stated choice experiments in transport planning environments the basis for the design of the experiment is the definition of trip utility. The trip utility expression is commonly the linear-in-parameters sum of various quantitative, observable trip cost, time and transfer attributes as well as qualitative attributes such as comfort and convenience. Attribute levels are prescribed for each attribute, and labels provide meaning to respondents such as Rands, minutes and the number of transfers. This hierarchical relationship is shown in Figure 1. Transport mode choice-related SP experiments consist of choice situations (commonly termed choice sets, C) that contain a number of mode alternatives (M), each with a number of modal attributes (A) (that do not have to be the same for each mode alternative), and with each attribute having a number of levels (L). The number of choice sets required for full factorial design (FD) for a labelled experiment is given by $C=L^{MA}$ (Street, et al., 2005; Bliemer & Rose, 2010) and $C=L^{M}$ for an unlabelled one. For example, a full factorial design for two labelled alternatives with three attributes each and with three attribute levels would require $3^{(2X3)}$ = 729 choice sets. The number would be 27 for an unlabelled design. Note that for level balance, the number of choice sets must be a factor of the number levels.

Even though it may constrain the design to be sub-optimal, attribute level balance is a desirable property that ensures that the levels for each attribute appear an equal number of times in the design, and hence avoids attribute level dominance by an alternative (Bliemer & Collins, 2015; Rose & Bliemer, 2009). The number of choice sets for a full factorial design is commonly too many for practical implementation, so fractional factorial designs (FFD) are used. These designs reduce the cognitive load place on respondents – the ideal number of choice sets presented to respondents is a complex mix of the number of alternatives, attributes, and levels (Chung, et al., 2011; Hensher et al., 2001). Instead of randomly choosing choice sets from a full factorial design, choice sets can be chosen to ensure that the attribute levels are orthogonal (i.e. to ensure no correlation between the levels of two attributes). Orthogonal designs can be sequential, i.e. orthogonality holds within each alternative, or simultaneous in which orthogonality also holds across alternatives. Sequential designs require fewer choice sets than simultaneous designs (ChoiceMetrics, 2018). The inherent assumption made for orthogonal FFD's is that there is no *a priori* information for the parameter estimates, and they are hence by definition set to zero. These parameters are the *b* values in the utility expressions shown below.

If the number of choice sets required by a FFD design is still too large for practical implementation, then a blocking strategy can be used, i.e. the allocation of choice sets into blocks with a respondent sample allocated to each block (Bliemer & Rose, 2010). When using a blocking strategy it is important to ensure each block is sampled with approximately the same number of responses per block (Bliemer & Collins, 2015). Response imbalance for a block design will result in attribute correlation in the estimated model. In the design estimation process, blocks are considered as another utility attribute across all alternatives with the number of attribute levels equal to the number of blocks.

For mode choice simulation in the context of urban peak period trip demand dominated by commuters, the main effects generic trip utility expression for a public transport mode is commonly the linear-in-parameters sum of the various trip time and cost attributes. The main effects are defined as the direct independent effect of each attribute on the response variable, i.e. choice, as opposed the interaction effects that are the effect on the response variable of combinations of two or more attributes (Hensher, et al., 2015). The case studies presented in this paper all have similar main effects utility expressions with generic parameters of the following form for individual *i* and public transport mode *j*:

$$U_{ij} = b_1*wtt_{ij} + b_2*wkt_{ij} + b_3*ivt_{ij} + b_4*c_{ij} + b_5*nt_{ij} + \varepsilon_i$$

where U_{ij} is the total trip (dis)utility; *wtt* is the waiting time attribute for mode *j*; *wkt* is the time attribute spent walking to the mode *j* stop or terminal; *ivt* is the time spent travelling in-vehicle on mode *j*; *c* is the trip cost or fare attribute for mode *j*; *nt* is the attribute defining the number of transfers to be made on the trip using mode *j*; ε_i is the unobserved error term associated with individual *i*; and $b_1, b_2 \ldots b_5$ are the parameters to be estimated. For

MNL models the error terms are assumed IID (independently and identically distributed) and drawn from a Type 1 generalised extreme value (GEV) distribution with mean zero and variance 1. For individual *i* and transport mode *j* (i.e. car mode) the utility equation is commonly as follows:

$$U_{ij} = b_3 * ivt_{ij} + b_4 * c_{ij} + \varepsilon_i$$

where *ivt* is the in-car travel time; *c* is the out-of-pocket trip cost that that is typically the petrol cost of the car trip; ε_i is the unobserved error term as before and b_3 and b_4 are the attribute parameters to be estimated. If the car trip has other out-of-pocket costs such as parking fees or tolls, these are added to the utility expression as separate attributes with their associated parameters. No toll or parking cost attributes were included in the case studies though there were several toll roads in operation for three of the case studies. In a choice experiment with these two modes, the alternative specific constant (ASC) for one mode would be normalised to zero. The ASC for the alternative mode would be included in its utility expression.

3. A DISCUSSION ON ORTHOGONALITY

Attribute level orthogonality is a mathematical constraint that requires all the attributes to be statistically independent of one another (Bliemer & Collins, 2015). Orthogonality requires that each possible pair of attribute levels appears an equal number of times in the design and in which the columns of the design display no correlation (ChoiceMetrics, 2018). Models estimated from non-orthogonal data are characterised by confounded attributes and manifest with the incorrect parameter signs (the equivalent of multicollinearity in linear multiple regression models), and high parameter mean standard errors.

However discrete choice models estimated from data collected with orthogonal designs seldom have zero attribute correlation (Bliemer & Rose, 2010) and this has called into question the requirement for orthogonal designs (Blamey, et al., 2002; Roman et al., 2011). Attribute level correlation occurs in non-linear models for several reasons including unequal spacing of the attribute levels; non-responses to some choice sets; and unequal representation in block samples if a blocking strategy has been used. Thus, while orthogonal designs are appropriate for linear models (such as multiple linear regression models) that meet these constraints, they are not suitable non-linear discrete choice models (ChoiceMetrics, 2018).

Evidence of the extent of the loss of orthogonality between an orthogonal design and the estimated MNL model is demonstrated by means of a case study for a choice model between two access modes (private car versus feeder bus mode) to a railway station. An orthogonal fractional factorial main effects experimental design was used for an SP survey between the two modes whose generic utility expressions were as follows:

$$U(car) = b_1 * (ivt)_{car} + b_2 * (cost)_{car} + b_3 * (parking\ cost)_{car}$$

and,

$$U(bus) = ASC_{bus} + b_1 (ivt)_{bus} + b_2 * (cost)_{bus} + b_4 * (wkt)_{bus} + b_5 * (wtt)_{bus} + b_6 * (safety)_{bus}$$

where *wkt* is the walking time; *wtt* is the waiting time; *ivt* is the in-vehicle travel time; *cost* is the petrol cost of the car trip and fare for the bus trip; *parking cost* is the park and ride fee at the railway station for the car mode; *safety* is personal safety dummy attribute for the walk part of the bus trip and b_1, b_2, ... b_6 are the parameters to be estimated. The *wkt,* *wtt, cost, parking cost and safety* attributes each had two levels and the *ivt* had three. Thirty six choice sets were required for an orthogonal fractional factorial design and a three block strategy was used with 12 choice sets per block. The response burden with 12 choice sets per respondent is suitable. Equal sampling for each of the blocks was achieved. The attribute correlation matrix of the MNL model is shown in Table 1. The interpretation of the correlation values is as follows (Profillides & Botzoris, 2019). For a value between 0 and ± 0.3, a weak correlation exists; between ±0.3 and ±0.6 a moderate correlation; between ±0.6 and ±0.8 a strong correlation exists, and for values greater than ±0.80 a very strong correlation.

Table 1: MNL Correlation Matrix for Car Versus E-Hail Access Mode Choice Model

Correlation Matrix	wkt	wtt	ivt	cost	parking	safety
wkt	1.00					
wtt	0.09	1.00				
ivt	-0.32	-0.21	1.00			
cost	0.00	0.01	-0.40	1.00		
parking	-0.05	-0.03	0.16	-0.60	1.00	
safety	0.01	0.00	0.06	-0.05	-0.05	1.00

4. EFFICIENT EXPERIMENTAL DESIGNS

Efficient experimental designs are a widely used alternative to fractional factorial designs. These designs release the orthogonality constraint and have as their focus the minimisation of parameter standard errors and hence the maximisation of the design efficiency. These designs make use of the fact that the square roots of the diagonal of the asymptotic variance covariance (AVC) matrix are the parameter asymptotic standard errors. When estimating a discrete choice model based on maximum likelihood, the second partial derivative of the likelihood function is the Hessian matrix. The negative of the Hessian is the Fischer Information Matrix which contains information relating to the rate of change of the likelihood function slope (i.e. the curvature) for each parameter. The negative inverse of the Fischer information matrix is the AVC matrix. The diagonal of the AVC matrix contains the parameter variances, and their square roots divided by the square root of the sample size are the parameter standard errors of the mean (Rose & Bliemer, 2009; ChoiceMetrics, 2018).

The AVC matrix can be determined analytically for estimation of an efficient design based on a single respondent requiring the analyst to specify the type of model to be estimated (multinomial logit (MNL), nested logit (NL) or random parameter logit (RPL)); the utility expressions for each alternative; the levels for each attribute; the number of choice sets (and whether a blocking strategy is to be used and if so the number of blocks); and *a priori* estimates of the attribute parameters (i.e. b_1, b_2, .. b_n) (Bliemer & Collins, 2015). These prior parameter values can be obtained from previous studies or from literature. Even if the sign of the parameters is known, the design can be improved (ChoiceMetrics, 2018). Designs from which MNL models are estimated require an estimate of the mean parameter prior value. If a random parameters logit (RPL) model is to be estimated then the distributions of the prior attribute parameters must be specified. For example, a normally distributed parameter requires the specification of the parameter mean and standard deviation.

More realism can be added by pivoting the alternative mode attribute levels off the observed attribute levels of the current mode, with the pivot levels specified by the analyst

(ChoiceMetrics, 2018). The perceived current mode attribute levels (i.e. anchor levels) must be solicited from the respondent for the alternative mode pivoting process that must be done in real time by the survey software. These designs thus use a combination of revealed preference (from the current mode observed attribute values) and stated preference (from the alternative mode pivoted attribute levels).

The design measure of efficiency is expressed as a measure of efficiency error which is to be minimised in the estimation process (Bliemer & Rose, 2010). The most widely used measure is the D-error which is the determinant of the AVC matrix. The design with the lowest D-error is termed the D-optimal design (Bliemer & Collins, 2015; Rose & Bliemer, 2009). Because the design with the lowest D-error is sometimes hard to find, the design with a sufficiently low D-error is used and this is termed the D-efficient design. Another error measure is the A-error and the design with the lowest A-error is called the A-optimal design. The A-error is the trace of the diagonal of the AVC matrix (i.e. the sum of the diagonal values). Hence the A-error only considers the variances and not the covariances of the AVC matrix.

An important characteristic of efficient designs is that the impact of the sample size can be investigated. Because estimates of the parameter asymptotic standard errors are the square root of the diagonals in the AVC matrix divided by the sample size, the standard errors decrease at the rate of $1/\sqrt{N}$, where N is the sample size. The standard errors thus decrease at a diminishing rate with increasing sample size. It is therefore more cost efficient to improve the accuracy of the prior parameter values than increasing the sample size (ChoiceMetrics, 2018).

5. METROPOLITAN EXPERIMENTAL DESIGN CASE STUDIES

The four metropolitan experimental design case studies are labelled Metro A, Metro B, Metro C and Metro D. Table 2 summarises the experimental designs of each for the car - BRT designs. The actual number of choice sets in the designs are shown, as well as the actual number of attribute levels that were used. The D-errors for these designs were estimated *ex post* and are also shown. Based on the actual design attributes and levels, the number of choice sets required for a full factorial design and an orthogonal fractional factorial design (ensuring no correlation between attributes) are shown together with their respective D-errors. Note that for Metro D two separate designs were evaluated. Metro D1 has car, train and BRT alternatives and was targeted at the middle-high income segment (gross monthly household income between R25 601 and R51 200 monthly). Metro D2 has car, bus and taxi alternatives and was aimed at the low-middle income segment (gross monthly household income between R3,201 and R25,600 per month).

Table 2 highlights several important issues. Firstly, the designs for Metro's A and B are the same. These designs used 20 choice sets per respondent, whereas a minimum of 24 are needed for an orthogonal fractional factorial design. To employ a design with 24 choice sets, a blocking strategy would be appropriate to reduce the cognitive burden on respondents, i.e. two blocks with 12 choice sets each. For Metro's A and B designs no car trip cost attribute was included in the utility expression which is an important oversight. For these metros the D-error is significantly improved by the orthogonal FFD over the value for the original design (0.009 versus 0.005, or a ratio of 1.80). This means that, on average, the asymptotic standard errors of the parameter estimates in the original design would be $\sqrt{1.8} = 1.34$, i.e. 34% higher than those estimated with the orthogonal FFD.

Table 2: Summary of Metro Experimental Designs for Multinomial Logit Models (MNL)

Experimental Criteria	Metro A	Metro B	Metro C	Metro D1	Metro D2
Modes	Car, BRT	Car, BRT	Car, BRT	Car, Train, BRT	Car, Bus, Taxi
Utility Attributes* (no. of attributes)	*Car:* ivt *BRT:* wtt, ivt, f, nt (4)	*Car:* ivt *BRT:* wtt, ivt, f, nt (4)	*Car:* ivt, f *BRT: wkt, w*tt, ivt, f, nt (4 per block)	*Car:* ivt, f *Train & BRT:* wtt, ivt, fare f, nt (4)	*Car:* ivt, f *Bus & Taxi:* wtt, ivt, f, nt (4)
Actual No. Choice Sets in Experiment Designs	20	20	18 (2 blocks of 9, block 1 with *wkt* & block 2 with *wtt*)	13	13
Actual No. of Attribute Levels in Experimental Designs	wtt = 4 ivt = 4 f = 4 nt = 3	wtt = 4 ivt = 4 f = 4 nt = 3	wkt = 3 wtt = 3 ivt = 3 f = 3 nt = 2	wtt = 4 ivt = 5 f = 7 nt = 3	wtt = 4 ivt = 7 f = 7 nt = 3
Actual Design D-error	0.0090	0.0090	n/a	0.0027	0.0021
Estimated no. of choice sets required for full factorial design (D-error)	768 (0.0002)	768 (0.0002)	486 (without pivot design) (0.030)	No valid design found – too many atts. & levels	No valid design found - too many atts. & levels
Estimated no. of choice sets required for orthogonal fractional factorial design (FFD) (D-error)	24 (0.0050)	24 (0.0050)	18 (without pivot design) (0.080)	27 (0.001)	No valid design found
Note: * wkt=walking time; wtt = waiting time; ivt=in-vehicle time; f=fare/cost; nt = no. of transfers					

For Metro C two separate designs were used even though they were stated to be two blocks of the same design. The first design (block 1) included the walking time (*wkt*) attribute to the BRT station (with three levels), and the second design (block 2) replaced the *wkt* with the waiting time (*wtt*) attribute (also with three levels) at the BRT station. It is not possible to use a blocking strategy with different utility expressions for the BRT mode in each block and the D-error of this design cannot thus be calculated. The orthogonal fractional factorial design for Metro C requires 18 choice sets assuming either *wkt* or *wkt* are included in the utility expression. If both *wkt* and *wtt* are included in the BRT utility expression with three levels each then an orthogonal fractional factorial design requires 36 choice sets (D-error is 0.042). A three-block strategy of 12 choice sets each would be appropriate in this case.

Both Metro D designs D1 and D2 have different modes with the same attributes but different attribute levels. For Metro D1 the original design consisted of three alternatives, four attributes and a range of attribute levels - from seven for trip fare (*f)* to three for the number of transfers (*nt*). The minimum required choice sets for an orthogonal fractional factorial design is 27, significantly more than the 13 choice sets used in the design. The D-error also reduces significantly from the original design value of 0.0027 to the orthogonal FFD value of 0.001 implying, on average, a 64% improvement in parameter standard errors with the orthogonal FFD design. For Metro D2, the number of attributes levels is high, i.e. seven for *ivt*, seven for *f*, four for *wtt* and three for the *nt* attribute. For this design there are too many attributes and levels for a full factorial or an orthogonal fractional factorial design to be identified.

Tables 3, 4 and 5 show the design correlation matrices for the original designs for Metros A, B and D1. It can be anticipated that due to the poor design efficiencies the MNL models estimated from these designs will be compromised. The correlations in Table 3 are high, and some appear to have the wrong signs. Some of the Metro D1 correlation values in Table 4 are high and may negatively influence the statistical significance of the MNL model. Several of the values for Metro D2 in Table 5 indicate strong correlation between attribute levels, and together with the low design efficiencies the MNL model estimated with this design is likely to be compromised.

Table 3: Attribute Correlation Matrix for Metros A and B

Attribute	car.ivt	brt.wtt	brt.ivt	brt.f	brt.nt
car.ivt	1				
brt.wtt	1.00	1			
brt.ivt	0.61	0.61	1		
brt.f	0.84	0.84	0.54	1	
brt.nt	0.83	0.83	0.87	0.64	1

Table 4: Attribute Correlation Matrix for Metro D1

Attribute	car.ivt	car.f	train.wtt	train.ivt	train.f	train.nt	brt.wtt	brt.ivt	brt.f	brt.nt
car.ivt	1									
car.f	0.00	1								
train.wtt	0.48	0.00	1							
train.ivt	-0.04	0.21	-0.08	1						
train.f	0.04	-0.66	0.28	-0.47	1					
train.nt	-0.29	0.12	0.29	-0.34	0.37	1				
brt.wtt	0.21	0.23	-0.07	-0.56	0.01	-0.25	1			
brt.ivt	0.03	-0.09	-0.22	-0.46	0.31	0.21	0.19	1		
brt.f	-0.14	-0.46	0.05	-0.31	0.59	0.05	-0.05	0.13	1	
brt.nt	-0.47	0.34	-0.66	0.07	-0.36	0.02	0.22	-0.11	-0.30	1

Table 5: Attribute Correlation Matrix for Metro D2

Attribute	car.ivt	car.f	bus.wtt	bus.ivt	bus.f	bus.nt	taxi.wtt	taxi.ivt	taxi.f	taxi.nt
car.ivt	1									
car.f	0.39	1								
bus.wtt	-0.05	-0.55	1							
bus.ivt	0.38	0.22	-0.16	1						
bus.f	-0.10	-0.32	0.71	0.10	1					
bus.nt	0.00	0.24	0.00	-0.14	0.31	1				
taxi.wtt	0.11	-0.13	0.15	0.07	-0.26	0.00	1			
taxi.ivt	0.02	-0.06	0.19	-0.56	-0.21	-0.30	0.34	1		
taxi.f	0.09	0.75	-0.44	0.43	-0.38	0.05	0.14	-0.28	1	
taxi.nt	0.42	0.17	0.03	-0.14	-0.11	-0.11	-0.33	0.28	-0.23	1

6. METRO MULTINOMIAL LOGIT (MNL) MODELS

Discrete choice MNL models were estimated from the SP data collected for each metro. Prior to model estimation the data sets were cleaned and non-traders and lexicographic errors were removed. The models are labelled MNL A, B, D1 and D2. For Metro C, two models were estimated, one for the utility expression containing the *wkt* attribute for BRT mode (MNL C1) and the other for utility expression that contains the *wtt* attribute for the BRT mode (MNL C2). The model outputs are summarised in Table 6.

Table 6: Metro Multinomial Logit (MNL) Model Output Summaries

Model Attribute	Metro A MNL A	Metro B MNL B	Metro C MNL C1	Metro C MNL C2	Metro D MNL D1	Metro D MNL D2
Log-Likelihood	-243	-475	-406	-445	-852	-663
Chi Square	6.0	29.8	195	239	160	7.7
Prob Chi^2 > Value	0.199	0.100	0.000	0.000	0.000	0.105
wkt Mean Coeff Std Error *t-value*	-	-	-0.016 0.021 *-0.75*	-	-	-
wtt Mean Coeff Std Error *t-value*	+0.020 0.016 *+1.22*	-0.038 0.010 *-3.77*	-	0.000 0.013 *0.00*	-0.083 0.009 *-9.38*	+0.011 0.125 *0.88*
ivt Mean Coeff Std Error *t-value*	+0.025 0.011 *+2.18*	-0.040 0.010 *-3.86*	-0.005 0.003 *-1.88*	-0.005 0.011 *-0.46*	-0.015 0.005 *-2.87*	-0.011 0.005 *-2.09*
f Mean Coeff Std Error *t-value*	+0.059 0.064 *0.92*	-0.099 0.077 *-1.29*	-0.163 0.014 *-11.36*	-0.163 0.013 *-12.5*	-0.055 0.010 *-5.31*	-0.022 0.01 *-2.18*
nt Mean Coeff Std Error *t-value*	+0.062 0.142 *0.44*	-0.291 0.177 *-1.65*	-0.320 0.180 *-1.78*	-0.326 0.176 *-1.85*	-0.388 0.166 *-2.34*	-0.045 0.077 *-0.59*
Car ASC Std Error *t-value*	+2.265 0.952 *2.38*	-2.44 1.183 *-2.07*	+0.625 0.255 *2.45*	+0.783 0.216 *3.63*	+0.512 0.403 *1.27*	+0.192 0.276 *0.69*
ASC Train Std Error *t-value*	-	-	-	-	-0.194 0.067 *-2.88*	-
ASC Bus Std Error *t-value*					-	0.105 0.174 *0.60*
Value of Travel Time (R/Hr)	n/a	n/a	R1.95/hr	n/a	R16.36/hr	R29.44/hr

Table 6 shows that there are deficiencies in all the models except MNL D1. The influence of the experimental designs is evident, i.e. parameters with the wrong signs (the time, cost and transfer parameter signs should all be negative reflecting the disutility of travel); some parameters are not significant (as indicated by their t-values within the critical range of ± 1.98 for the 95% confidence interval); the associated high standard errors; and in some cases the probability of Chi^2 exceeding a value larger than 0.05 (for the 95% confidence interval). These Chi^2 values indicate that the attribute parameters are not significantly different to zero when considered collectively in the utility expression.

Other model characteristics that raise concerns are as follows. The *wtt:ivt* ratio for models MNL A, MNL B and MNL C2 are less than 1.0. This implies that travellers would prefer to wait than to travel which is implausible. This ratio for model MNL D1 is 5.5, which is significantly higher than benchmarked values of between 1.8 and 3.0 (Wardman, 2011). It is not appropriate to calculate the value of travel time for MNL A, MNL B and MNL C2 as the fare parameters are not significant for MNL A and MNL B, and the in-vehicle time parameter is not significant for MNL C2. The value of travel time for MNL C1 is R1.85 per hour, which is unrealistically low, while the value of MNL D2 of R29.44/hour appears plausible but cannot be relied on due to the poor overall validity of the model. The MNL D1 value of R16.36/hour seems plausible given that MNL D1 is a statistically significant model. Overall, the model results for Metros A, B, C1, C2 and D2 cannot be used for forecasting prediction. Model D1 should be used with caution, as the *wtt:ivt* ratio is far outside benchmarked limits.

The attribute correlation matrices for three of the models are shown in Table 7 (MNL A), Table 8 (MNL C1) and Table 9 (MNL D1). The range of correlation values is significant, with higher correlations present for models MNL A and MNL D1. For model A the correlation between Fare (f) and ASC Car is high (0.94). This is likely due to the omission of a trip cost attribute for cars in the utility expression. The MNL A correlation values derived from the models are generally lower than those derived from the experimental design (Table 4). There is a similar pattern for MNL D1 as shown in Table 9.

Table 7: Correlation Matrix for Model MNL A

Attribute	WTT	IVT	F	NT	ASC Car
WTT	1				
IVT	0.14	1			
F	0.31	0.38	1		
NT	-0.29	0.15	0.26	1	
ASC Car	0.40	0.60	0.94	0.37	1

Table 8: Correlation Matrix Model MNL C1

Attribute	WKT	IVT	F	NT	ASC Car
WKT	1				
IVT	0.03	1			
F	0.02	0.07	1		
NT	-0.02	0.05	0.06	1	
ASC Car	0.81	0.05	-0.12	0.43	1

Table 9: Correlation Matrix for Model MNL D1

Attribute	WTT	IVT	F	NT	Car ASC	Train ASC
WTT	1					
IVT	0.02	1				
F	-0.04	-0.24	1			
NT	0.01	-0.26	-0.22	1		
Car ASC	0.33	-0.19	-0.81	0.38	1	
Train ASC	-0.18	-0.07	0.08	-0.02	-0.03	1

7. ESTIMATION OF EFFICIENT DESIGNS

D-efficient experimental designs have been prepared for three of the case studies, i.e. for Metro's A, C1 and D1. For Metro's A and D1, MNL models were derived. For Metro C1, a random parameters logit (RPL) model was estimated. For each design the prior parameter values have been derived from MNL D1 (see Table 6), with the *wtt* parameter value adjusted to a value of two times the *ivt* value. The mean prior parameter values (and standard deviations) are therefore *wtt* = -0.030 (0.009); *ivt* = -0.015 (0.005); f = -0.055; and nt = -0.384 (0.166). Note that the ASC attributes are not included in the designs but will be estimated in the MNL and RPL models. The d-efficient design parameters for each Metro are summarised in Table 10. For a discussion of the impact of error in the prior estimates see Bliemer & Collins (2015) and Walker et al (2018).

The experimental design for an RPL model requires the prior parameters to be randomised according to the distribution intended in the RPL. The *wtt, ivt* and *nt* parameters were randomised normal and the cost parameter *f* was not randomised to overcome the problematic issue of the ratio of two distributions when estimating willingness-to-pay measures (Train, 2009). 100 Halton draws have been used for the randomised draws. The panel effects must also be taken into consideration (ChoiceMetrics, 2018), i.e. the effect of repeated observations by individual respondents. This design is also a pivot design, with the BRT in-vehicle travel time and fare pivoted off the (observed) car values for this metro. Note that the *ivt* pivot levels are defined in percentage terms and the *wtt, f* and *nt* levels variation are defined as values as shown in Table 10. For the C1 design, the observed car travel time and cost were the mean observed values for this metro, viz. 30 minutes and R25 respectively.

Several important issues should be kept in mind when specifying the designs. Firstly, the attribute levels should be realistic in the context of the experiment and also to the respondent. The model estimated from the surveys will only be valid for the level ranges specified in the design. Secondly, for this reason it is recommended that wider level ranges are specified with fewer levels (Bliemer & Collins, 2015). These so-called end-point designs define two extreme levels (that must still be realistic) that usually translate into smaller asymptotic standard errors. However non-linearities cannot be estimated with two level designs. Thirdly, there should be attribute level balance in the design, i.e. no alternatives should dominate others as discussed earlier. If dominance does occur (i.e. the design is unbalanced) the choice situation does not provide information for estimating the parameters and the design will not be efficient. Hence the experimental design needs to trade off the needs of the choice context as well as the constraints for efficient designs.

Table 10: Efficient Design Parameters for Metro A, C1 and D1

Design Parameter	Metro A	Metro C1 (Pivot Design)	Metro D1
Alternatives	Car, BRT	Car, BRT	Car, Train, BRT
Attributes & Levels	*wtt* [5, 15] ivt_{car} [15,25,30] f_{car} [15,25,40] ivt_{brt} [20,35,50] f_{brt} [10,15,20] *nt* [0, 1]	*wtt* [5, 10, 15] *ivt** [-10%, 0%, 20%] *f*** [-15, -10, 10] *nt* [0, 1]	*wtt* [5, 10, 15] *ivt* [30, 50, 70] *f* [10, 20, 30] *nt* [0, 1, 2]
No. of Choice Sets	36 (3 blocks of 12)	18 (2 blocks of 9)	15
*Reference value for car ivt = 30 mins for homogeneous pivot design; **Reference value for car cost is R25 for homogeneous pivot design. *wtt* and *ivt* are in minutes, *f* in Rands and *nt* number of transfers.			

The efficient designs are shown in Tables 11, 12 and 13 for Metros A, C1 and D1 respectively. The correlation matrix for each design is also shown (as measured by the Pearson Product Moment). The D-error for the equivalent orthogonal fractional factorial designs for Metro A and Metro D1 are also shown. In the output the B-estimate is the minimum sample size required for overall parameter significance. The S-estimate and the Sp estimates are the minimum sample sizes required for individual parameter significance. The Sp t-ratios are the estimated t-values that will be achieved with a single respondent. The S-estimate and Sp estimate should only be considered as the lower bound sample size requirements when comparing alternative designs.

The main observations from these designs are that the efficient designs are all improvements over the equivalent orthogonal designs except for the pivot design for Metro C1 from which an RPL model will be developed. An equivalent orthogonal pivot design cannot be estimated for an RPL or MNL, but an orthogonal design without pivoting results in an improved D-error. The suggested minimum sample sizes for Metro A is 85 respondents, 87 for Metro C1 and 71 for Metro D1. Based on the prior parameter distribution defined for the *ivt* attribute for Metro C1 (i.e. normal with mean -0.015 and standard deviation 0.005), the suggested minimum sample size for this attribute (Sp estimate) is 329 respondents. The equivalent pivoted design for an MNL model has a lower D-error of 0.015 and requires a minimum suggested sample size of 90 respondents. This result highlights the requirement for larger sample sizes required for RPL models and also emphasizes the need for further research into the application of efficient designs for RPL models. Some attribute level correlations are high for Metro's A and D1, but generally lower for Metro C1.

Table 11: Metro A Efficient Design for MNL

D error	0.00292				For equivalent orthogonal FFD the D-error is 0.003408 and A-error is 0.097801. Hence on average, efficient design mean parameter standard errors will be 8% more accurate.		
A error	0.09188						
B estimate	85.2						
S estimate	9.5						
Prior	***b1 ivt***	***b3 f***	***b2 wtt***	***b4 nt***			
Fixed prior value	-0.015	-0.055	-0.03	-0.384			
Sp estimates	6.05	0.82	6.45	9.51			
Sp t-ratios	0.80	2.17	0.77	0.64			
Design							
Choice Set No.	car.ivtc	car.fc	brt.wtt	brt.ivtb	brt.fb	brt.nt	Block
1	15	15	5	50	15	1	3
2	25	25	15	35	15	0	3
3	30	25	15	20	15	0	3
4	25	25	15	20	15	0	2
5	15	40	15	50	10	0	1
6	15	40	5	50	10	0	1
7	25	40	15	35	10	0	3
8	25	15	5	50	20	1	1
9	15	15	15	50	20	0	3
10	30	40	5	20	10	1	2
11	30	25	15	20	10	0	3
12	25	40	5	35	10	1	3
13	30	15	15	20	15	1	1
14	25	40	15	20	10	0	1
15	15	25	5	50	20	1	2
16	30	15	15	20	20	0	1
17	30	15	15	35	20	0	3
18	15	25	15	50	15	0	3
19	25	15	15	35	15	0	1
20	15	40	5	50	15	0	2
21	15	25	5	50	15	0	2
22	30	15	15	20	20	1	2
23	15	40	5	50	10	1	1
24	25	15	15	35	20	0	1
25	25	40	5	35	15	1	2
26	15	15	5	35	20	1	2
27	25	15	5	35	20	1	2
28	15	25	15	50	20	0	1
29	30	25	15	20	20	1	3
30	30	40	5	35	10	1	2
31	15	25	5	50	15	1	3
32	30	25	15	20	15	1	1
33	30	15	5	35	20	1	2
34	25	40	5	20	10	1	3
35	30	25	5	20	10	1	1
36	25	40	5	35	10	0	2
Correlations (Pearson Product Moment)							
Attribute	car.ivtc	car.fc	brt.wtt	brt.ivtb	brt.fb	brt.nt	Block
car.ivtc	1.00	-0.13	0.27	-0.85	-0.03	0.18	0.00
car.fc	-0.13	1.00	-0.30	0.05	-0.79	-0.05	0.02
brt.wtt	0.27	-0.30	1.00	-0.34	0.20	-0.56	0.00
brt.ivtb	-0.85	0.05	-0.34	1.00	0.13	-0.14	0.00
brt.fb	-0.03	-0.79	0.20	0.13	1.00	0.07	0.00
brt.nt	0.18	-0.05	-0.56	-0.14	0.07	1.00	0.00
Block	0.00	0.02	0.00	0.00	0.00	0.00	1.00

Table 12: Metro C1 Pivot Efficient Design for RPL with Panel Effects (100 Halton Draws)

D error	0.039			Note: Efficient Design for MNL model (no pivot) has D-error=0.016 and A-error =0.189.			
A error	0.384						
B estimate	87.0						
S estimate	329						
Prior	***b2 ivt***	***b3 f***	***b1 wtt***	***b4 nt***			
Fixed prior value	-0.015	-0.055	-0.030	-0.384			
Sp estimates	329.4	2.9	17.3	19.6			
Sp t-ratios	0.11	1.16	0.47	0.44			
Design							
Choice situation	**car.ivt**	**car.f**	**brt.wtt***	**brt.ivt (pivot)***	**brt.f (pivot)**	**brt.nt***	**Block**
1	30	25	15	0%	-10	0	1
2	30	25	15	20%	-10	0	2
3	30	25	5	-10%	-15	1	2
4	30	25	10	0%	-10	1	1
5	30	25	5	0%	-10	1	2
6	30	25	10	0%	-10	0	2
7	30	25	10	-10%	-15	0	1
8	30	25	10	20%	-15	1	1
9	30	25	5	0%	10	1	1
10	30	25	15	-10%	10	0	2
11	30	25	5	20%	10	1	2
12	30	25	5	-10%	-15	0	2
13	30	25	15	20%	10	0	2
14	30	25	10	20%	-15	0	1
15	30	25	15	0%	-15	1	2
16	30	25	10	20%	-10	0	1
17	30	25	5	-10%	10	1	1
18	30	25	15	-10%	10	1	1
Correlations (Pearson Product Moment)							
Attribute	**car.ivt**	**car.f**	**brt.wtt**	**brt.ivt**	**brt.f**	**brt.nt**	**Block**
car.ivt	!	!	!	!	!	!	!
car.f	!	!	!	!	!	!	!
brt.wtt	!	!	1.00	0.16	0.03	-0.41	0.00
brt.ivt	!	!	0.16	1.00	-0.06	-0.18	0.00
brt.f	!	!	0.03	-0.06	1.00	0.21	0.00
brt.nt	!	!	-0.41	-0.18	0.21	1.00	-0.11
Block	!	!	0.00	0.00	0.00	-0.11	1.00

wtt and *ivt* are in minutes, *f* in Rands and *nt* number of transfers.
*wtt, ivt and nt are randomised normal, f is fixed

Table 13: Metro D1 Efficient Design

D-error	0.004				**Note:** Equivalent orthogonal design requires 27 choice sets (3 blocks x 9 choice sets) with D-error = 0.002 and A-error=0.027. Minimum sample size is 100 respondents.					
A-error	0.038									
B estimate	70.8									
S estimate	11.0									
Prior Parameters	***b2 ivt***	***b3 f***	***b1 wtt***	***b4 nt***						
Fixed prior value	-0.015	-0.055	-0.03	-0.384						
Sp estimates	5.7	1.8	11.0	3.8						
Sp t-ratios	0.82	1.46	0.59	1.00						
Design										
Choice Set No.	**car.ivt**	**car.f**	**train.wtt**	**train.ivt**	**train.f**	**train.nt**	**brt.wtt**	**brt.ivt**	**brt.f**	**brt.nt**
1	30	30	10	70	10	2	5	70	20	2
2	70	30	5	50	30	1	10	30	10	0
3	70	20	5	50	10	2	10	30	10	0
4	70	20	15	50	30	1	5	30	20	2
5	30	10	10	50	20	0	15	30	20	0
6	70	10	15	30	30	2	5	30	30	2
7	70	30	5	50	10	2	15	70	10	0
8	30	10	15	50	20	1	5	70	20	1
9	70	30	5	50	10	2	15	50	30	1
10	50	20	10	50	20	0	10	50	20	1
11	30	30	15	30	30	0	15	70	20	0
12	50	10	10	30	20	1	5	30	30	1
13	30	10	10	50	20	1	5	70	20	1
14	50	30	15	70	10	0	10	50	10	2
15	70	30	15	30	30	0	10	70	10	2
Correlations (Pearson Product Moment)										
Attribute	**car.ivt**	**car.f**	**train.wtt**	**train.ivt**	**train.f**	**train.nt**	**brt.wtt**	**brt.ivt**	**brt.f**	**brt.nt**
car.ivt	1.00									
car.f	0.32	1.00								
train.wtt	-0.31	-0.21	1.00							
train.ivt	-0.21	0.28	-0.23	1.00						
train.f	0.09	-0.18	0.51	-0.66	1.00					
train.nt	0.37	0.00	-0.51	0.13	-0.40	1.00				
brt.wtt	0.12	0.49	-0.38	-0.04	-0.20	-0.20	1.00			
brt.ivt	-0.42	0.34	0.19	0.12	-0.18	-0.09	0.09	1.00		
brt.f	-0.18	-0.50	0.15	-0.34	0.11	0.23	-0.26	-0.21	1.00	
brt.nt	0.09	0.00	0.61	0.13	0.10	0.00	-0.61	0.09	0.23	1.00
wtt and *ivt* are in minutes, *f* in Rands and *nt* number of transfers.										

8. CONCLUSIONS

This paper has demonstrated that the experimental designs used in four South African metros for the development of mode choice simulation models were inadequate, not meeting several design requirements associated with orthogonal fractional factorial designs. These include the number of choice sets required for attribute level balance for orthogonality and the appropriate blocking strategies. The consequences were that in all instances the MNL models estimated from the data collected from the experiments were not statistically significant, leaving a large gap in the understanding of urban car commuter mode choice trip behaviour. The willingness to pay for travel time savings as measured by the value of travel time is hence still uncertain for urban South African car commuters. The analysis provided evidence that orthogonal designs are not orthogonal when executed, with a case study showing high attribute correlation in some instances.

The paper demonstrated that main effects orthogonal fractional factorial designs based on the utility expressions described in this paper would improve the D-error over the original designs. Efficient designs would further improve the D-error, and the use of efficient designs for MNL models with the prior parameter values determined in this analysis is recommended. While the use of pivot-based efficient designs is recommended for MNL models, further research is required for their application in RPL models in the South African context. The research should focus on the identification and specification of the appropriate parameter distributions.

9. REFERENCES

Blamey, R, Bennett, J, Louviere, J & Morrison, J, 2002. Attribute Causality in Environmental Choice Modelling. *Environmental and Resource Economics,* 23:167-186.

Bliemer, M & Collins, A, 2015. Chapter 6: Experimental Design and Choice Experiments. In: D Hensher, J Rose & W Greene, eds. *Applied Choice Analysis.* Cambridge: Cambridge University Press, pp. 189-318.

Bliemer, M & Rose, J, 2010. Chapter 6: Serial Choice Conjoint Analysis for Estimating Discrete Choice Models. In: S. Hess & A. Daly, eds. *Choice Modelling: The State of the Art and the State of Practice.* Bingley: Emerald Publishing, pp. 139-161.

ChoiceMetrics, 2018. *Ngene 1.2 User Manual & Reference Guide.* 1.2 ed. Sydney: Choice Metrics.

Chung, C, Boyer, T & Han, S, 2011. How Many Choice Sets and Alternatives are Optimal? Consistency in Choice Experiments. *Agribusiness,* 27(1):114-125.

Hensher, D, Rose, J & Greene, W, 2015. *Applied Choice Analysis.* 2nd ed. Cambridge: Cambridge University Press.

Hensher, D, Stopher, P & Louviere, J, 2001. An Exploratory Analysis of the Effect of Numbers of Choice Sets in Designed Choice Experiments: An Airline Choice Application. *Journal of Air Transport Management,* 7(1):373-379.

Keuol, R & Burger, D, 2018. *Design of Experiments: Statistical Principles of Research Design and Analysis.* 1st ed. Andover: Cengage Learning EMEA.

Profillides, V & Botzoris, G, 2019. *Modelling of Transport Demand: Analyzing, Calculating, and Forecasting Transport Demand.* 1 ed. Amsterdam: Elsevier.

Puckett, S & Rose, J, 2010. Chapter 7: Oberved Efficiency of a Do-Optinal Design in an Interactive Agency Choice Experiment. In: S. Hess & A. Daly, eds. *Choice Modelling: The State of the Art and State of Practice.* Bingley: Emerald Publishing, pp. 163-192.

Roman, C, Martin, J, Espino, R & Arencibia, A, 2011. *Working Paper 2/2011: Efficient Versus Non-Efficient Stated Choice Designs: A Comparison in a Mode Choice Context,* Las Palmas de Gran Canaria: CREI (Universidad de Las Palmas de Gran Canaria.

Rose, J & Bliemer, M, 2009. Constructing Efficient Stated Choice Experimental Designs. *Transport Reviews,* 29(5):587-617.

Street, D, Burgess, L & Louviere, J, 2005. Quick and Easy Choice Sets: Contructing Optimal and Nearly Optimal Stated Choice Experiments. *International Journal of Research in Marketing,* 22:459-470.

Train, K, 2009. *Discrete Choice Methods with Simulation.* Second Edition ed. Cambridge: Cambridge University Press.

Walker, J et al., 2018. D-Efficient or Deficient? A Robustness Analysis of Stated Choice Experimental Designs. *Theory and Decision,* 84:215-238.

Wardman, M, 2011. Public Transport Values of Time. *Transport Policy,* 11:363-377.