

Research Article

Comparing the State-of-the-Art Efficient Stated Choice Designs Based on Empirical Analysis

Li Tang,¹ Xia Luo,¹ Yang Cheng,² Fei Yang,¹ and Bin Ran²

¹ School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China

² Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA

Correspondence should be addressed to Li Tang; tanglitraffic@gmail.com

Received 16 November 2013; Accepted 12 January 2014; Published 20 February 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Li Tang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The stated choice (SC) experiment has been generally regarded as an effective method for behavior analysis. Among all the SC experimental design methods, the orthogonal design has been most widely used since it is easy to understand and construct. However, in recent years, a stream of research has put emphasis on the so-called efficient experimental designs rather than keeping the orthogonality of the experiment, as the former is capable of producing more efficient data in the sense that more reliable parameter estimates can be achieved with an equal or lower sample size. This paper provides two state-of-the-art methods called optimal orthogonal choice (OOC) and *D*-efficient design. More statistically efficient data is expected to be obtained by either maximizing attribute level differences, or minimizing the *D*-error, a statistic corresponding to the asymptotic variance-covariance (AVC) matrix of the discrete choice model, when using these two methods, respectively. Since comparison and validation in the field of these methods are rarely seen, an empirical study is presented. *D*-error is chosen as the measure of efficiency. The result shows that both OOC and *D*-efficient design are more efficient. At last, strength and weakness of orthogonal, OOC, and *D*-efficient design are summarized.

1. Introduction

Abundant and accurate data is the foundation of study. To date, emerging technologies are largely introduced in data mining and processing [1]. In [2, 3], how to obtain high-quality data with the high-tech in the field of Intelligent Transportation System is deeply discussed. However, the importance of refining data collection technique has not raised researchers' attention in behavior analysis area until last decades, such as travel mode choice and safety in [4, 5]. The purpose of conducting choice experiments is to collect data that can be used to estimate the independent influence of attributes on observed choices. There are two paradigms of choice data: revealed preference (RP) and stated choice (SC) data. Typically, in RP surveys, respondents are asked to recall information about his/her last choice, including alternatives and attribute levels available in real market. Differently, SC experiments present sampled respondents with a number of different hypothetical choice situations, each consisting of a universal but finite set of alternatives defined on a

number of attribute dimensions. Thus, SC data collected on 300 respondents, each of whom is asked to make 8 choices produces a total of 2400 choice observations while RP experiment on the same sample size of respondents collects only 300 observed choices.

SC experiment has been widely used because it can observe choices on alternatives which do not exist in the current market. So analysts are able to predict, for example, the share rate of a newly introduced transportation mode. Another reason of its popularity lies in the ability to provide variability in attributes in a relatively small sample size compared with RP experiment, with which better estimation of influence of each attribute on choice can be achieved. Usually, respondents in a SC experiment will be faced with some "selected" choice situations, considering that making choices among all the possible combinations of attribute levels is too many to accomplish for a single respondent. Thus, how analysts distribute the levels of the design attributes in an experiment plays a big role. It may impact upon not only whether or not an independent assessment of contribution

of each attribute to the choices observed can be determined, but also the ability of the experiment to detect statistical relationships that may exist within the data.

Historically, researchers have relied on orthogonal experimental designs, in which the attributes of the experiment are statistically independent by forcing them to be orthogonal [6]. As such, orthogonal designs theoretically allow for an independent determination of each attribute's influence upon the observed choices. To generate an orthogonal design, usually but not necessarily, the first step is to generate a full factorial design, a design which contains all possible attribute level combinations. Mathematically, a full factorial design will produce $\prod_{k=1}^k L_k$ choice situations, where L_k is the number of levels assigned to attribute k (e.g., a design with 4 attributes, two with 2 attribute levels, one with 3 levels, and one with 4 levels will produce a full factorial design with $2 \times 2 \times 3 \times 4 = 48$ choice situations). The second step is to take a subset of choice situations from the full factorial design, which is known as fractional factorial designs. Randomized, cyclical, Bayesian, and fold over procedures are the common approaches used to generate fractional factorial design in [7–10]. It is noteworthy that, in either a full factorial or a fractional factorial design, orthogonality is kept between two random attributes. The only difference between them is that orthogonality will not be kept in terms of interaction effects (i.e., the influence of two or more attribute columns multiplied together) in fractional factorial design. Here are some simple rules to check whether a fractional factorial design is orthogonal or not:

- (i) Every level of every attribute appears at same times.
- (ii) All possible attribute level combinations of random two attributes appear at same times.

While orthogonal design has long been used in practice, in [11, 12], there is a stream of researchers in recent years doubted the importance of orthogonality in SC data when it is used to estimate discrete choice model, not to mention whether orthogonality can be kept in reality. Orthogonality is important in linear models since it avoids multicollinearity problem and also minimized variance-covariance matrix of the estimated model, in which way standard errors of parameter estimates are also minimized. Unfortunately, discrete choice model is nonlinear; thus the derivation of its parameters' variance-covariance matrix is very different from the way in linear models. Seeing from that, keeping orthogonality of the parameters has little to do with minimizing their standard errors.

Acknowledgment of this fact has led researchers to transfer their efforts to obtain experimental designs that minimize the asymptotic variance-covariance (AVC) matrix of discrete choice models and provide more reliable parameter estimates with an equal or lower sample size. Such designs are called efficient designs. To date, most research has been focused on developing methods to generate efficient designs; comparison and validation of these methods in practice are rarely seen. This paper provides two state-of-the-art methods called optimal orthogonal choice (OOC) and D -efficient design. Their performance on both theoretical efficiency and

TABLE 1: Optimal orthogonal choice design for 2 alternatives with 3 binary attributes.

Choice situation	Alternative 1			Alternative 2		
	X_1	X_2	X_3	X_1	X_2	X_3
1	0	0	0	1	1	1
2	0	1	1	1	0	0
3	1	0	1	0	1	0
4	1	1	0	0	0	1

practical use are compared with the conventional orthogonal design.

The remainder of this paper is organized as follows. In Section 2, OOC and D -efficient design are introduced. In Section 3, SC experiment design using orthogonal, OOC, and D -efficient methods separately is generated. The result in terms of D -errors are presented in Section 4, along with the summarized strength and weakness.

2. Efficient Experimental Design Methods

2.1. Optimal Orthogonal Choice Design. Considering the popularity and convenience of orthogonal design for analyzers in practice, there is a stream of researchers in [13, 14] who kept on exploring improved SC design method maintaining orthogonality, which is called optimal orthogonal choice design. The essential idea of OOC design is to maximize the differences of attribute levels across alternatives, so that the parameters can be estimated in the largest extent of variety of attribute levels as well as independently. The basic process of generating an OOC design is as follows.

Step 1. Generate a fractional factorial orthogonal design for alternative 1. N represents the number of choice situations of the design.

Step 2. Choose some systematic changes to get the allocation of attribute levels in alternative 2 from alternative 1. Systematic changes are certain rules to decide how the attribute levels change from alternative 1 and will be discussed in later context.

Step 3. Choose another systematic change to get the allocation of attribute levels in alternative 3 from alternative 1.

Step 4. Keep doing this until all the alternatives are determined.

It will be much easier to understand this method by starting with a binary attribute level design. Again, we assume that L_k is the number of levels assigned to generic attribute k for alternative j , represented by $0, 1, \dots, L_k - 1$. In a design for 2 alternatives and 3 attributes each with 2 levels, an orthogonal design in 4 choice situations for alternative 1 can be firstly generated. Then 0's and 1's in alternative 1 are interchanged in alternative 2. Thus the attribute levels of each attribute are forced to be different across alternatives. The result is shown in Table 1.

TABLE 2: Optimal orthogonal choice design for 3 alternatives with 3 binary attributes.

Choice situation	Alternative 1			Alternative 2			Alternative 3		
	X_1	X_2	X_3	X_1	X_2	X_3	X_1	X_2	X_3
1	0	0	0	1	1	0	0	0	1
2	0	1	1	1	0	1	0	1	0
3	1	0	1	0	1	1	1	0	0
4	1	1	0	0	0	0	1	1	1

To generate OOC design for more alternatives, it is necessary to introduce S_k to represent the largest number of different pairs appeared between alternatives for a specific attribute. The equation of S_k is shown as follows, where J stands for the number of alternatives in choice set:

$$S_k = \begin{cases} \frac{(J^2 - 1)}{4}, & l_k = 2, J \text{ odd,} \\ \frac{J^2}{4}, & l_k = 2, J \text{ even,} \\ \frac{(J^2 - (l_k x^2 + 2xy + y))}{2}, & 2 < l_k \leq J, \\ \frac{J(J-1)}{2}, & l_k \geq J. \end{cases} \quad (1)$$

In an example of OOC design for 3 alternatives and 3 attributes each with 2 levels shown in Table 1, we can get $S_k = (J^2 - 1)/4 = 2$ for all the attributes. For instance, in the first choice situation (000, 110, and 001), the attribute levels differ twice for each attribute (i.e., for attribute X_1 , the levels are 010, creating 3 pairs (01, 10, and 00) in which two of them (01, 10) are different).

We can see from Table 1 that the distribution of attribute levels in alternative 2 is obtained by interchanging 0's and 1's in X_1 and X_2 in alternative 1 and that in alternative 3 is obtained by interchanging 0's and 1's in X_3 in alternative 1. These systematic changes can be also described as adding a generator in alternative 1 to get alternative 2 and adding another generator to get alternative 3. The addition is performed in modulo arithmetic according to the number of levels for a specific attribute. Here in the example, $L_k = 2$ for all attributes, thus when a generator 110 is added to the choice situations in alternative 1 in modulo 2 arithmetic like this: $000 + 110 \equiv 110$, $011 + 110 \equiv 101$, and so on, alternative 2 is obtained. Alternative 3 is generated in the same way by adding a generator 001 to alternative 1. Notice that the generators added to alternatives must have a value of $S_k = 2$ (i.e., generators (000,110, and 001) used in Table 2 can meet the requirement while another generators (000, 100, and 010) cannot not).

Designs for any choice set size with any number of attributes each having any number of levels can be generated in similar way. However, a big limitation of OOC design is that it can only generate designs for generic attributes according to the principle of this method. How alternative-specific attributes distribute across alternatives is rarely discussed

in the literature. Another shortcoming of OOC design is that it may produce a lot of unreasonable combinations by forcing maximum level differences in attributes across alternatives. Answers from responders who have to make decision in such choice situations may not reflect their actual choice statements since the "proper alternative" may not be contained in the questionnaire.

2.2. D-Efficient Design. While people who raise OOC method keep on improving design with remaining orthogonality, another stream of researchers goes straight forward to increasing the statistical efficiency of the design by minimizing the elements of the asymptotic variance-covariance (AVC) matrix of discrete choice models. The AVC matrix can be obtained by taking the negative inverse of the expected second derivatives of the log-likelihood function of the model proofed in [15]. To interpret the process of calculating the AVC matrix and the measure of efficiency of a design, here we briefly introduce the most well-known multinomial logit model.

Assume an individual faced with alternative $j = 1, 2, \dots, J$ in choice situation $n = 1, 2, \dots, N$. The utility of an individual for alternative j in choice situation n can be expressed as

$$U_{jn} = V_{jn} + \varepsilon_{jn}, \quad (2)$$

where V_{jn} represents observed part of utility for each alternative j in choice situation n . It is assumed to be a linear additive function of several attributes with corresponding weights. These weights are unknown parameters to be estimated and can be divided into two categories: generic parameters and alternative-specific parameters. The generic parameters and alternative-specific parameters can be denoted by β_k^* , $k = 1, \dots, K^*$, and β_{jk} , $k = 1, \dots, K_j$, respectively, with their associated attribute levels x_{jkn}^* and x_{jkn} for each choice situation n . Thus, the total number of parameters to be estimated equal to $\bar{K} = K^* + \sum_{j=1}^J K_j$. V_{jn} can be expressed as

$$V_{jn} = \sum_{k=1}^{K^*} \beta_k^* x_{jkn}^* + \sum_{k=1}^{K_j} \beta_{jk} x_{jkn}, \quad \forall j = 1, \dots, J, \quad \forall n = 1, \dots, N, \quad (3)$$

where ε_{jn} is the unobserved component, which is independently and identically extreme value type one distributed. The probability P_{jn} that an individual chooses alternative j in choice situation n becomes

$$P_{jn} = \frac{\exp(V_{jn})}{\sum_{j=1}^J \exp(V_{jn})}, \quad \forall j = 1, \dots, J, \quad \forall n = 1, \dots, N. \quad (4)$$

Considering that the most popular way to estimate parameters is maximum likelihood estimation, the log-likelihood function of parameters for a single respondent can be expressed as

$$L(\beta^*, \beta) = \sum_{n=1}^N \sum_{j=1}^J y_{jn} \log P_{jn} \\ = \sum_{n=1}^N \left[\sum_{j=1}^J y_{jn} \left(\sum_{k=1}^{K^*} \beta_k^* x_{jkn}^* + \sum_{k=1}^{K_j} \beta_{jk} x_{jkn} \right) - \log \left(\sum_{j=1}^J \exp \left(\sum_{k=1}^{K^*} \beta_k^* x_{jkn}^* + \sum_{k=1}^{K_j} \beta_{jk} x_{jkn} \right) \right) \right], \quad (5)$$

where y represents the binary outcome of all choice situations. While alternative j is chosen in choice situation n , y_{jn} equals one; otherwise it is zero. Then the AVC matrix can be expressed as the second derivative of the log-likelihood function as follows:

$$\frac{\partial^2 L(\beta^*, \beta)}{\partial \beta_{k_1}^* \partial \beta_{k_2}^*} = - \sum_{n=1}^N \sum_{j=1}^J x_{jk_1 n}^* P_{jn} \left(x_{jk_2 n}^* - \sum_{i=1}^J x_{ik_2 n}^* P_{in} \right), \quad (6) \\ \forall k_1, k_2 = 1, \dots, K^*,$$

$$\frac{\partial^2 L(\beta^*, \beta)}{\partial \beta_{j_1 k_1} \partial \beta_{j_2 k_2}^*} = - \sum_{n=1}^N x_{j_1 k_1 n} P_{j_1 n} \left(x_{j_2 k_2 n}^* - \sum_{i=1}^J x_{ik_2 n}^* P_{in} \right), \quad (7) \\ \forall j_1 = 1, \dots, J, \quad k_1 = 1, \dots, K_{j_1}, \quad k_2 = 1, \dots, K^*,$$

$$\frac{\partial^2 L(\beta^*, \beta)}{\partial \beta_{j_1 k_1} \partial \beta_{j_2 k_2}} = \begin{cases} \sum_{n=1}^N x_{j_1 k_1 n} x_{j_2 k_2 n} P_{j_1 n} P_{j_2 n}, & \text{if } j_1 \neq j_2; \\ - \sum_{n=1}^N x_{j_1 k_1 n} x_{j_2 k_2 n} P_{j_1 n} (1 - P_{j_2 n}), & \text{if } j_1 = j_2. \end{cases} \\ \forall j_1 = 1, \dots, J, \quad k_i = 1, \dots, K_{j_i}. \quad (8)$$

Equations (6)–(8) represent functions that allow generic and alternative-specific parameters. In the case where only generic parameters exist, only (6) remains, and when there are only alternative-specific parameters, (8) remains. In addition, if there are M identical respondents, these second derivatives are multiplied by M .

Let $(\bar{\beta}^*, \bar{\beta})$ denote the true values of the parameters. The Fisher information matrix I is defined as the expected values of the second derivative of the log-likelihood function:

$$I(\bar{\beta}^*, \bar{\beta}) = M \cdot \frac{\partial^2 L(\bar{\beta}^*, \bar{\beta})}{\partial \beta \partial \beta'}. \quad (9)$$

Hence, the AVC matrix can be expressed as a $\bar{K} \times \bar{K}$ matrix that is equal to the negative inverse of the Fisher information matrix [16]:

$$\Omega = -[I(\bar{\beta}^*, \bar{\beta})]^{-1} = -\frac{1}{M} \left[\frac{\partial^2 L(\bar{\beta}^*, \bar{\beta})}{\partial \beta \partial \beta'} \right]^{-1}. \quad (10)$$

Rather than working with each element within the AVC matrix directly, a preferred measure within the literature is D -error, calculated by taking the determinant of the AVC matrix and scaling this value by the number of parameters \bar{K} . It is common to assume that a single respondent (i.e., $M = 1$) represents all respondents, an assumption consistent with the multinomial logit (MNL) model form [8, 17]. Designs that aim at minimizing D -error are called D -efficient designs.

Since the calculation of D -error involves the values of parameters, approaches to determine D -error have been improved in recent years. In our empirical study later, we use D_p -error as the statistic to measure the efficiency of experimental designs. To calculate D_p -error, nonzero priors are needed. It can be expressed as follows:

$$D_p\text{-error} = - \left[\det \frac{\partial^2 L(\bar{\beta}^*, \bar{\beta})}{\partial \beta \partial \beta'} \right]^{-1/\bar{K}}. \quad (11)$$

According to the D -efficient method, we can figure out that, for a given sample size, attempts to minimize D -error statistic will directly lead to the minimization of the AVC matrix. Meanwhile, by taking the square root of the diagonal elements (including M) of the AVC matrix, the minimization of asymptotic standard errors is achieved. Thus the asymptotic standard error of the parameter estimates will diminish in terms of statistical significance from each additional respondent added to a survey.

Figure 1(a) reveals that the standard error decreases by increasing sample size at the beginning of a given design X^I . Exceeding a certain limit, enlarging sample size has little impact on demising standard error. On the other side, Figure 1(b) shows that investing a more efficient design X^{II} can lead to larger decreases in standard error. In other words, a smaller sample size may be satisfied at a certain level of standard error when using efficient designs rather than common ones.

3. Empirical Study

In this section, three experimental designs which used different methods (orthogonal, OOC, and D -efficient) are generated. They will be used to obtain trip mode choice data on a corridor connecting two large business districts in Chengdu. Shawan conference and exhibition center and Jinsha station are chosen as the origin and destination for the survey. Three alternatives are involved: car, taxi, and bus. A typical multinomial logit model in transportation will be formulated and serve as the basis of most of the analyses in

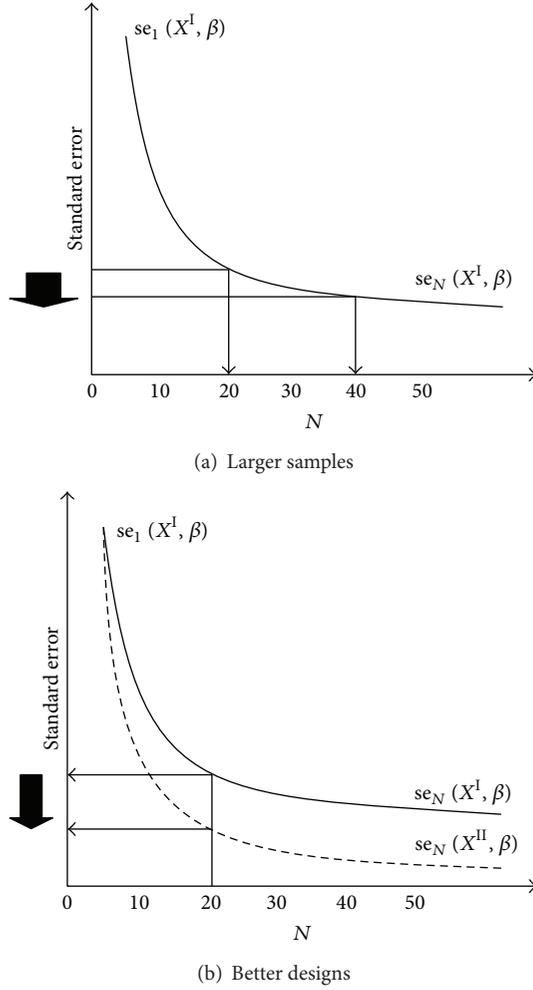


FIGURE 1: Comparison of investing in larger sample sizes versus more efficient designs.

the subsequent section. The observed part of utility of every alternative is expressed as follows:

$$\begin{aligned}
 V^{\text{car}} &= \beta_0^{\text{car}} + \beta_1 \text{TT}^{\text{car}} + \beta_2 \text{TC}^{\text{car}}, \\
 V^{\text{taxi}} &= \beta_0^{\text{taxi}} + \beta_1 \text{TT}^{\text{taxi}} + \beta_2 \text{TC}^{\text{taxi}}, \\
 V^{\text{bus}} &= \beta_1 \text{TT}^{\text{bus}} + \beta_2 \text{TC}^{\text{bus}},
 \end{aligned} \quad (12)$$

where TT represents travel time. TC represents travel cost. For car users, TC equals the fuel cost. For taxi users, TC equals the money paid for the trip. For bus users, TC equals the ticket price. Seeing from (12), parameters for TT and TC are generic across three alternatives. Thus, four parameters are going to be estimated in total (two of them are alternative-specific constant, which has nothing to do with any attribute). The attribute levels and prior information about parameters are given in Table 3 based on previous study results as well as to preserve realistic estimates for the private and public transport alternatives.

In order to obtain better estimation of parameters, three levels are set for each attribute for maximum variation as

TABLE 3: Prior parameter values and attribute levels for case study.

(a)					
Prior parameter values					
β_0^{car}	β_0^{taxi}	β_1	β_2		
1	-0.6	-0.45	-0.9		
(b)					
Attribute level					
Car		Taxi		Bus	
TT ^{car} (min)	TC ^{car} (RMB)	TT ^{taxi} (min)	TC ^{taxi} (RMB)	TT ^{bus} (min)	TC ^{bus} (RMB)
15	4	15	16	25	1
20	5	20	18	30	2
30	7	30	22	38	3

much as possible. The values of TT^{car} , TT^{taxi} , and TT^{bus} are measured in free, normal, and congested traffic flow. The values of TC^{car} are calculated as the kilometers between the origin and destination by the consumed oil price under 7.3 RMB/liter, 7.7 RMB/liter, and 8 RMB/liter. The values of TC^{taxi} are measured in free, normal, and congested traffic flow. The value of TC^{bus} is based on the current price and plus/minus 1 RMB. The number of choice situations (i.e., 18) is selected such that attribute level balance can be achieved. Obviously, this number is too large for a single respondent. Thus, we introduce a block variable to divide the design into smaller parts (i.e., here we block the design into three parts so that six choice situations are provided to a single respondent). Each block is not orthogonal by itself, but in combination with other blocks. Attribute level balance is maintained as much as possible in each block.

We generate three different (attribute level balanced) designs with 18 choice situations assuming the above MNL model, using the software Ngene 1.1.1. The design results are shown in Table 4 as well as D -error value for each design.

4. Results and Discussion

As expected, the two efficient designs produce lower D -error value (0.118313 and 0.114612 for the OOC design and D -efficient design, resp.), while orthogonal design produces higher D -error value (0.194724), seeing from Table 4. The D -error of the orthogonal design is 1.64 times greater than the D -error value of the OOC design and 1.69 times greater than D -efficient design. This suggests that, on average, the asymptotic standard errors of the parameter estimates using the orthogonal design will be 1.28 to 1.31 times larger than the efficient designs. Clearly, the efficient designs are able to provide more reliable parameter estimates than orthogonal design.

On the other hand, comparing the two efficient designs, D -efficient outperforms OOC in terms of the D -error value (0.114612 versus 0.118313). Furthermore, since the OOC method can only generate designs with generic attributes, the use of it is largely limited. In a word, with high statistical

TABLE 4: Experimental designs for empirical study.

Choice situation	Car		Taxi		Bus		Block
	TT ^{car} (min)	TC ^{car} (RMB)	TT ^{taxi} (min)	TC ^{taxi} (RMB)	TT ^{bus} (min)	TC ^{bus} (RMB)	
Orthogonal design for MNL model (D -error = 0.194724)							
1	15	4	15	16	25	1	1
2	20	5	20	18	30	2	1
3	30	7	30	22	38	3	1
4	15	5	20	22	38	1	1
5	20	7	30	16	25	2	1
6	30	4	15	18	30	3	1
7	30	7	20	18	25	1	2
8	15	4	30	22	30	2	2
9	20	5	15	16	38	3	2
10	20	7	15	22	30	1	2
11	30	4	20	16	38	2	2
12	15	5	30	18	25	3	2
13	30	5	30	16	30	1	3
14	15	7	15	18	38	2	3
15	20	4	20	22	25	3	3
16	20	4	30	18	38	1	3
17	30	5	15	22	25	2	3
18	15	7	20	16	30	3	3
Orthogonal optimal design for MNL model (D -error = 0.118313)							
1	15	4	20	18	38	3	1
2	30	7	15	16	30	2	1
3	15	5	20	22	38	1	1
4	20	4	30	18	25	3	1
5	30	5	15	22	30	1	1
6	20	7	30	16	25	2	1
7	20	7	30	16	25	2	2
8	30	5	15	22	30	1	2
9	15	4	20	18	38	3	2
10	20	4	30	18	25	3	2
11	15	5	20	22	38	1	2
12	30	7	15	16	30	2	2
13	15	7	20	16	38	2	3
14	20	5	30	22	25	1	3
15	30	4	15	18	30	3	3
16	30	4	15	18	30	3	3
17	20	5	30	22	25	1	3
18	15	7	20	16	38	2	3
D -efficient design for MNL model (D -error = 0.114612)							
1	30	5	20	18	25	1	1
2	20	7	15	16	38	2	1
3	30	4	15	22	25	3	1
4	15	5	30	16	30	3	1
5	15	5	30	16	38	2	1
6	30	5	20	22	25	1	1
7	20	4	15	22	30	2	2
8	15	7	30	22	38	1	2
9	15	7	20	22	38	1	2
10	30	5	15	16	30	3	2

TABLE 4: Continued.

Choice situation	Car		Taxi		Bus		Block
	TT ^{car} (min)	TC ^{car} (RMB)	TT ^{taxi} (min)	TC ^{taxi} (RMB)	TT ^{bus} (min)	TC ^{bus} (RMB)	
11	15	5	30	18	38	2	2
12	20	4	20	18	25	3	2
13	20	7	30	16	25	1	3
14	20	4	20	18	30	2	3
15	30	7	15	16	30	3	3
16	15	7	30	18	38	1	3
17	20	4	20	18	30	3	3
18	30	4	15	22	25	2	3

TABLE 5: Comparison of orthogonal, OOC, and *D*-efficient design.

Method	Advantage	Disadvantage
Orthogonal	(i) It is the most widely used method and easy to construct or obtain (ii) There are no correlations between attribute levels; thus it allows for an independent estimation of the influence of each attribute on choice	(i) There are too many choice situations/questions for a single respondent (ii) Orthogonally it is hard to maintain in actual design: subsets replicated unevenly, introducing sociodemographic variable and allocation bias of the implausible choice situation (iii) It may contain “useless” choice situations
Optimal orthogonal choice	(i) Attribute level differences are maximized (ii) Choice situations will be reduced as well as attaining the design’s orthogonality	(i) It can only generate designs for generic attributes; the rules for setting up alternative-specific attributes are not clear right now (ii) Unreasonable combinations of attribute levels may appear; thus the “real choice” of respondents is hard to capture
<i>D</i> -efficient	(i) The smaller the asymptotic standard errors achieved, the smaller the width of the confidence intervals observed around the parameters estimates will be (ii) <i>t</i> -Radios will be maximized thus producing more reliable study results and analyst is able to minimize the sample size	(i) In general not orthogonal (not that important) (ii) Advanced knowledge of the parameter estimates is needed (iii) It needs more computation power

efficiency and wide applicability, *D*-efficient design achieves the best performance.

Further, most research focused on developing one of these experimental design methods by far. Though there may be a few of discussions about strength and shortcomings of every method separately, comparisons are rarely found in theory or practice area [12, 18, 19]. Here, we conclude the advantages and disadvantages of these three methods in Table 5. Also, a popularity rate is given as a reference of their applications in the field.

5. Conclusion

The SC experiment has been generally regarded as an effective method for discrete choice analysis, especially for newly introduced alternatives. The high cost on survey forces researchers to find more efficient design methods to obtain better estimation on parameters instead of investing a larger sample size. Though orthogonal design has been used as the

major experimental design method, orthogonality is not that important in the nonlinear discrete choice models. In this paper, we provide two state-of-the-art efficient designs: OOC and *D*-efficient design. By comparing orthogonal, OOC, and *D*-efficient design in both theory and practice, we find that efficient designs are more capable of producing more efficient data in the sense that more reliable parameter estimates can be achieved with an equal or lower sample size. The generation process requires the assumption of prior parameter estimates and model structure to construct AVC matrix. The result suggests a move away from orthogonal designs for SC experiments towards *D*-efficient designs, which make relative discrete choice models being more fit with such data.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This study is supported by the 2014 Doctoral Innovation Funds of Southwest Jiaotong University, the Fundamental Research Funds for the Central Universities (no. A0920502051307-03), Specialized Research Fund for the Doctoral Program of Higher Education (no. 20130184110020), National Natural Science Foundation of China (no. 51178403), and Sichuan Province Science and Technology Support Project (no. 2011FZ0050).

References

- [1] H. Tan, G. Feng, J. Feng, W. Wang, Y. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transportation Research Part C*, vol. 28, pp. 15–27, 2013.
- [2] H. Tan, J. Feng, G. Feng, W. Wang, and Y.-J. Zhang, "Traffic volume data outlier recovery via tensor model," *Mathematical Problems in Engineering*, Article ID 164810, 8 pages, 2013.
- [3] H. Tana, B. Chenga, J. Fenga, G. Fenga, W. Wang, and Y. J. Zhangb, "Low-n-rank tensor recovery based on multi-linear augmented lagrange multiplier method," *Neurocomputing*, vol. 119, pp. 144–152.
- [4] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [5] W. Wang, X. Jiang, S. Xia, and Q. Cao, "Incident tree model and incident tree analysis method for quantified risk assessment: an in-depth accident study in traffic operation," *Safety Science*, vol. 48, no. 10, pp. 1248–1262, 2010.
- [6] J. J. Louviere, D. A. Hensher, and J. D. Swait, *Stated Choice Methods: Analysis and Applications*, Cambridge University Press, Cambridge, UK, 2000.
- [7] P. Galilea and J. D. D. Ortúzar, "Valuing noise level reductions in a residential location context," *Transportation Research D*, vol. 10, no. 4, pp. 305–322, 2005.
- [8] D. S. Bunch, J. J. Louviere, and D. Anderson, "A comparison of Experimental Design Strategies for Choice-based Conjoint Analysis with Generic-attribute Multinomial Logit Models," Working Paper, Graduate School of Management, University of California, Rickey Davis, Ala, USA., 1994.
- [9] J. J. Louviere and G. Woodworth, "Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data," *Journal of Marketing Research*, vol. 20, pp. 350–367, 1983.
- [10] J. Huber and K. Zwerina, "The importance of utility balance in efficient choice designs," *Journal of Marketing Research*, vol. 33, no. 3, pp. 307–317, 1996.
- [11] J. M. Rose and M. C. J. Bliemer, "Constructing efficient stated choice experimental design," ITLS Working Paper ITLS-WP-05-07, 2005.
- [12] M. C. J. Bliemer, J. M. Rose, and D. A. Hensher, "Efficient stated choice experiments for estimating nested logit models," *Transportation Research B*, vol. 43, no. 1, pp. 19–35, 2009.
- [13] D. J. Street and L. Burgess, "Optimal and near-optimal pairs for the estimation of effects in 2-level choice experiments," *Journal of Statistical Planning and Inference*, vol. 118, no. 1-2, pp. 185–199, 2004.
- [14] L. Burgess and D. J. Street, "Optimal designs for choice experiments with asymmetric attributes," *Journal of Statistical Planning and Inference*, vol. 134, no. 1, pp. 288–301, 2005.
- [15] D. McFadden, "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, P. Zarembka, Ed., pp. 105–142, Academic Press, New York, NY, USA, 1974.
- [16] D. McFadden, "The choice theory approach to market research," *Marketing Science*, vol. 5, no. 4, pp. 275–297, 1986.
- [17] J. M. Rose and M. C. J. Bliemer, "Stated preference experimental design strategies," in *Transport Modeling Handbooks in Transport*, D. A. Hensher and K. Button, Eds., vol. 1, chapter 8, Elsevier Science, Oxford, UK, 2nd edition, 2007.
- [18] J. M. Rose, M. C. J. Bliemer, D. A. Hensher, and A. T. Collins, "Designing efficient stated choice experiments in the presence of reference alternatives," *Transportation Research B*, vol. 42, no. 4, pp. 395–406, 2008.
- [19] D. J. Street, L. Burgess, and J. J. Louviere, "Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments," *International Journal of Research in Marketing*, vol. 22, no. 4, pp. 459–470, 2005.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

