

## Research Article

# An Empirical Study of Parameter Estimation for Stated Preference Experimental Design

Fei Yang,<sup>1,2</sup> Lin Chen,<sup>1</sup> Yang Cheng,<sup>3</sup> Xia Luo,<sup>1,2</sup> and Bin Ran<sup>4</sup>

<sup>1</sup> School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China

<sup>2</sup> Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Si Pai Lou No. 2, Nanjing 210096, China

<sup>3</sup> Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA

<sup>4</sup> School of Transportation, Southeast University, No. 2 Si Pai Lou, Nanjing, Jiangsu 210096, China

Correspondence should be addressed to Lin Chen; [chenlin\\_traffic@163.com](mailto:chenlin_traffic@163.com)

Received 12 July 2014; Accepted 15 August 2014; Published 31 August 2014

Academic Editor: Huimin Niu

Copyright © 2014 Fei Yang 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 preference experimental design can affect the reliability of the parameters estimation in discrete choice model. Some scholars have proposed some new experimental designs, such as D-efficient, Bayesian D-efficient. But insufficient empirical research has been conducted on the effectiveness of these new designs and there has been little comparative analysis of the new designs against the traditional designs. In this paper, a new metro connecting Chengdu and its satellite cities is taken as the research subject to demonstrate the validity of the D-efficient and Bayesian D-efficient design. Comparisons between these new designs and orthogonal design were made by the fit of model and standard deviation of parameters estimation; then the best model result is obtained to analyze the travel choice behavior. The results indicate that Bayesian D-efficient design works better than D-efficient design. Some of the variables can affect significantly the choice behavior of people, including the waiting time and arrival time. The D-efficient and Bayesian D-efficient design for MNL can acquire reliability result in ML model, but the ML model cannot develop the theory advantages of these two designs. Finally, the metro can handle over 40% passengers flow if the metro will be operated in the future.

## 1. Introduction

The stated preference has become the primary way of acquiring travelers' preference data regarding different transport services. In the survey process, respondents choose their preferred travel mode in some hypothetical scenarios. The designs of these hypothetical scenarios depend on the experimental design. Thus, the stated preference questionnaire design is based on the experimental design method [1]. Currently, the widely used experimental designs for stated preference questionnaire are the random fractional factorial design and orthogonal design [2–4]. Although random fractional factorial design can effectively reduce the number of combinations, the method is likely to produce statistically inefficient or suboptimal designs. To overcome the deficiencies associated with the random fractional factorial design, the orthogonal design was introduced into the stated preference questionnaire design. The orthogonal design was

aimed at ensuring independence among the attributes to avoid parameter estimation bias, which tends to result from multicollinearity [5]. However, a major shortcoming of the orthogonal design exists: due to the restriction of orthogonality among these attributes, the orthogonal design cannot determine the number of experimental combinations according to the research need. If too many alternative attributes and attribute levels are selected, the number of experimental combinations obtained by the orthogonal design becomes very large. Moreover, some scholars have questioned the effect of the reliability of estimating the model parameters based on the orthogonal design [6]. Some researchers have recognized that the experimental design would significantly impact the parameter estimation. They construct theoretical basis for some experimental designs in order to improve the reliability of parameter estimation results [7–9]. The most common ones are the D-efficient and Bayesian D-efficient design. Both designs have the same theoretical basis. The only

difference between the two designs is determined by whether the values of the prior parameters are fixed. The effectiveness of these new experimental designs was demonstrated by simulations data, but little empirical research has been performed in actual environments. Bliemer validated the theoretical advantages of the Bayesian D-efficient design in an actual environment, showing that this design can be effectively applied in practice [10]. Louviere believed that the D-efficient design may cause significant errors in the estimation of parameter variance [11]. This notion is pending validation in follow-up studies. Although there has been few comparative analyses of the traditional designs [4, 12], there has been little comparative analysis of the new designs against the traditional designs. Although insufficient empirical research has been conducted on the effectiveness of these new designs, a few researchers have started using these new experimental designs in relevant studies [13, 14].

There are some limitations in the existing papers on experimental investigations of the stated preference experimental design. In these works, the predominantly used model is the multinomial logit (MNL) model [8, 10]. However, the MNL model exists with a significant defect. That is, the traveler preferences heterogeneity and the correlation in attributes are not considered. However, in the experimental combinations obtained by the D-efficient and Bayesian D-efficient design, there exist certain correlations in the alternative attributes, and the Bayesian D-efficient design even considers the differences in individual preferences. There lacked adequate studies about whether these designs which were generated for the MNL model could work well in the advanced models, for example, the mixed logit (ML) model. The ML model takes into account the traveler preference heterogeneity and the correlation in attributes. Some researchers illustrated this problem through simulation data. They noted that the design generated specifically for the MNL model tends to perform well when more advanced models are used [15]. However, whether this conclusion holds with actual data still requires further validation.

Therefore, in this paper, the new metro between the city of Chengdu and one of its satellite cities is taken as the research subject. From an empirical point of view, comparative analysis of the new designs against the orthogonal design was made for an estimation of the model parameters; the validities of the D-efficient and Bayesian D-efficient design are empirically verified under different sample sizes. We would demonstrate whether these designs generated specially for the MNL model could work well in ML model based on the actual data.

## 2. The Experimental Design for the Stated Preference Questionnaire

*2.1. Orthogonal Design.* As a traditional stated preference experimental design, orthogonal design has been widely used in discrete choice model. The orthogonal design can reduce the number of combinations. By selecting mutually orthogonal combinations, the orthogonal design avoids the model parameter estimation errors that are caused by the correlation among the attributes [5]. The combinations that

are independent, nonrepeatable, and representative were taken as the questionnaire. For example, suppose there are two available alternatives, A and B, for the travelers and that each alternative contains the attributes  $K$  and  $L$  (e.g., waiting time and cost). Although the attributes of each alternative are the same, the level of each corresponding attribute is different. According to the principle of orthogonal design, all of the orthogonal and nonrepeated combinations are selected as the questionnaire content. Each column is orthogonal with respect to the other columns in the combination that is ultimately obtained; that is,  $\text{corr}(\vec{K}_A, \vec{K}_B) = 0$ ,  $\text{corr}(\vec{K}_A, \vec{L}_B) = 0$ ,  $\text{corr}(\vec{K}_A, \vec{L}_A) = 0$ , and  $\text{corr}(\vec{L}_B, \vec{K}_B) = 0$ . However, a major shortcoming of orthogonal design exists: because the orthogonal design is restricted by the condition of orthogonality, it cannot select the number of combinations based on the research need. The number of combinations obtained by the orthogonal design is related to the number of alternatives and attribute levels. Hence, orthogonal design lacks flexibility when choosing the number of combinations.

*2.2. D-Efficient Design.* The theoretical basis of the D-efficient design is to obtain the minimum value of the determinant of the asymptotic-covariance matrix (AVC) of the model. The essence of this design resides in the goal of obtaining minimum values of the estimated standard deviations of the model parameters, thereby obtaining more reliable parameters estimation results. This design explicitly considers the importance of alternative attributes to ensure that the combinations can present more trade-off information in the selection process, with the goal of maximizing the respondents' preference information [16]. Typically, the determinant of the AVC is expressed using D-error. The AVC represents the reliability of the model parameter estimation results. Therefore, the D-error value reflects the efficiency of the experimental design. Smaller values of D-error indicate higher reliability of the estimated parameter results.

According to the theoretical basis of the D-efficient design, when using this method, first, a prior AVC or prior parameter estimation results must be obtained to determine the experimental combination iteratively, using the minimum D-error value. Usually, we identify the previously obtained model AVC or model parameter estimate value as the prior information. How this prior information is obtained before the stated preference questionnaire is designed remains an issue that must be overcome when adopting this method. To overcome this issue, researchers collect data using a pilot survey. The pilot survey data are used to estimate the model parameter. The parameter estimation values are used as the prior information for questionnaire design. The mathematical derivation of the D-efficient design, based on related theory, is as follows:

$$\text{D-error} = \det(\Omega)^{1/K}, \quad (1)$$

where  $\Omega$  is the AVC matrix for the discrete choice model and  $K$  is the number of model parameters.

The AVC matrix is the inverse matrix of the Hessian matrix of the maximum likelihood function for the discrete choice model, as shown below:

$$\Omega = I^{-1}. \quad (2)$$

The specific form of the Hessian matrix is given as follows:

$$I = \begin{bmatrix} \frac{\partial^2 L}{\partial \beta_1^2} & \cdots & \frac{\partial^2 L}{\partial \beta_1 \beta_{k_2}} \\ \vdots & \frac{\partial^2 L}{\partial \beta_{k_1} \beta_{k_2}} & \vdots \\ \frac{\partial^2 L}{\partial \beta_1 \beta_{k_2}} & \cdots & \frac{\partial^2 L}{\partial \beta_{k_2}^2} \end{bmatrix}, \quad (3)$$

where each element in the Hessian matrix is calculated by the following equation:

$$\frac{\partial^2 L(X, \beta)}{\partial \beta_{j_1 k_1} \partial \beta_{j_2 k_2}} = \begin{cases} -\sum_{n=1}^N X_{nj_1 k_1} X_{nj_2 k_2} P_{nj_1} (1 - P_{nj_2}), & j_1 = j_2 \\ \sum_{n=1}^N X_{nj_1 k_1} X_{nj_2 k_2} P_{nj_1} P_{nj_2}, & j_1 \neq j_2, \end{cases} \quad (4)$$

where  $L(X, \beta)$  is the maximum likelihood function for the discrete choice model and  $\beta_{jk}$  is the parameter value of the  $k$ th attribute for the  $j$ th alternative.

The maximum likelihood function for the discrete choice model is shown below:

$$L(X, \beta) = \sum_{n=1}^N \sum_{i=1}^I y_{ni} \log P_{ni}, \quad (5)$$

where  $y_{ni}$  represents the choice result of the respondents,

$$y_{ni} = \begin{cases} 1, & \text{the travel } n \text{ chose the alternative } i \\ 0, & \text{other,} \end{cases} \quad (6)$$

and  $P_{ni}$  represents the choice probability of alternative  $i$ . The formula used to calculate  $P_{ni}$  is as follows:

$$P_{ni} = \frac{\exp(V_{ni})}{\sum_{j=1}^J \exp(V_{nj})}, \quad (7)$$

where  $J$  is the alternative set.

$V_{ni}$  represents the utility function of alternative  $i$  for traveler  $n$ . It is calculated as follows:

$$V_{ni} = \alpha_{ni} + \beta_{ni1} X_{ni1} + \beta_{ni2} X_{ni2} + \cdots + \beta_{nik} X_{nik}, \quad (8)$$

where  $X_{nik}$  represents the  $k$ th attribute for the  $i$ th alternative and  $\beta_{nik}$  is the unknown parameter in the utility function for the  $k$ th attribute for the  $i$ th alternative.

It can be observed that the D-efficient design is actually the inverse of the process that is used for parameter estimation in the discrete choice model. The theoretical bases for Bayesian D-efficient and D-efficient are identical. The only difference is that, in the Bayesian D-efficient design process, the prior values of the attribute parameters obey a certain distribution [17]. This type of experimental design takes into account the differences in travelers' preference.

TABLE 1: The definition table of the alternative attribute levels.

Alternative	Attributes	Level <sub>1</sub>	Level <sub>2</sub>	Level <sub>3</sub>
Bus	In-vehicle time (min)	60	90	120
	Cost (¥)	2	4	
	Waiting time (min)	5	10	
	Arrival time (min)	5	10	
	Off-vehicle time (min)	5	10	
Coach	In-vehicle time (min)	40	60	
	Cost (¥)	6	8	10
	Waiting time (min)	15		
	Arrival time (min)	20		
	Off-vehicle time (min)	15		
Metro	In-vehicle time (min)	30	40	
	Cost (¥)	4	6	
	Waiting time (min)	3		
	Arrival time (min)	10	30	
	Off-vehicle time (min)	10		
Taxi	In-vehicle time (min)	40	60	
	Cost (¥)	50	70	
	Waiting time (min)	5		
Car	In-vehicle time (min)	40	60	
	Cost (¥)	20	40	

### 3. Method Validation and Empirical Analysis: A Case Study

In this paper, the new metro between Chengdu and Longquan (a satellite city) is used as the research subject. Based on the MNL model, the influence of different experimental design on model parameter estimation is verified. Additionally, whether the design generated specifically for the MNL model is able to obtain reliable parameter results in a more complex ML model is verified. At present, the existing modes of transportation between the cities include bus, taxis, private cars, and coach. With the continued development of the two cities, Chengdu is ready to introduce the metro to meet the diverse travel needs of its residents.

**3.1. Design of the Stated Preference Questionnaire.** The intention of this stated preference questionnaire is to investigate the travelers' choice preference when introducing a subway system. Therefore, the alternative set in the stated preference questionnaire design process contains five alternatives, namely, the metro, bus, taxis, private cars, and coach. The alternative attributes include "arrival time (the time of arriving at the station)," "waiting time," "cost," "in-vehicle time," and "off-vehicle time (travel time from getting off the vehicle to the destination)." Finally, the alternative attributes and attribute levels are formed as presented in Table 1.

**3.1.1. Orthogonal Stated Preference Questionnaire.** Orthogonal design was used to design the stated preference questionnaire. According to the alternative attributes and the number of attributes shown in Table 1, there are a total of 32 mutually orthogonal combinations which were obtained by Negene

(a kind of experimental design software). The number of combinations obtained from the orthogonal design is related to the alternative attributes and the number of attribute levels. Therefore, this design cannot select the number of experimental combinations according to the research need. If we only used a part of combinations that maybe leads to a lack of orthogonality, the lack of orthogonality will cause the multicollinearity among attributes, thereby affecting the accuracy of the model results. Therefore, all 32 combinations were brought into the stated preference questionnaire. Using a block experimental design, under the premise of ensuring orthogonality, the 32 combinations were divided into 8 blocks. Each respondent attended to one block. Finally, the investigation instructions, filling instructions, attribute combinations, and personal attributes should be added; the complete orthogonal stated preference questionnaire was formed.

**3.1.2. D-Efficient and Bayesian D-Efficient Stated Preference Questionnaires.** The theoretical bases for the D-efficient and Bayesian D-efficient design are the same. The only difference lies in the form of prior parameter. The value of prior parameter is fixed in the D-efficient design. The prior values in the Bayesian D-efficient design are subject to normal distribution. As stated earlier, before the D-efficient design can be applied, the researcher must obtain prior information. The MNL model parameter estimation values obtained from the pilot survey data are set as the prior information for the D-efficient and Bayesian D-efficient design. The stated preference questionnaire used for the pilot survey also is designed by the orthogonal design. Using the pilot survey, 192 observations were obtained. The MNL model parameter values obtained from the pilot survey data are shown in Table 2. The prior values of the attribute parameters for D-efficient design are equal to the parameter estimation results in the MNL model. However, some differences exist in the process of setting the prior values of attribute for the Bayesian D-efficient design. The prior values for the significant attribute parameters are all subject to the  $N(\beta, \sigma)$  normal distribution, and the nonattribute parameters are all subject to the  $N(0, \sigma)$  normal distribution; the standard deviation ( $\sigma$ ) is equal to the standard error of the parameter estimations.

According to the prior values for the parameters shown in Table 2, the stated preference questionnaire based on the D-efficient design was completed. For comparison with the orthogonal design, the number of combinations also was determined to be 32. Ultimately, the obtained D-error was 0.227. Similarly, according to the blocking experimental design, the 32 combinations were divided into eight blocks, and each respondent was assigned to one block. Finally, after incorporating the additional investigation instructions, filling instructions, attribute combinations, and personal attributes, a complete D-efficient stated preference questionnaire was formed.

The prior values of attribute parameter for the Bayesian D-efficient design are subject to normal distribution, but the values of the mean and variance are different. According

to the prior values provided in Table 2, the number of combinations was also defined as 32. The corresponding D-error was 0.225. Similarly, using the blocking method, the 32 combinations were divided into eight blocks. The Bayesian D-efficient stated preference questionnaire was formed.

**3.2. Data Analysis.** In this paper, three types of questionnaires were utilized. Face-to-face survey was conducted. The preference data for travelers in different scenarios were obtained. Meanwhile, the respondents also were asked to provide personal information regarding their ages, income, and so forth. The overall survey sample size was 960. The gender rate between male and female is 1.03. The average age of overall samples is 36. The monthly average income of overall samples is 3556 (¥). By establishing four age intervals, each questionnaire was able to collect the age information from the surveyed sample. Of four groups, the 25–50 age range occupied the highest proportion. Similarly, five income ranges were established to evaluate the income levels of the respondents. The corresponding survey sample attributes of each questionnaire are shown in Table 3. According to Table 3, there are no significant differences between the general characteristics of the three survey samples, indicating that the general characteristics of the survey sample should not significantly influence the model parameter estimation results. Among the three groups, the rate of car ownership is relatively high and is a result of the fact that the rate of vehicle possession in Chengdu ranks first in the whole nation.

## 4. Analysis of the Results of Different Models

**4.1. Error Term Variance of the Different Experimental Design Methods.** First, we analyze whether a difference exists in the variance of the error term for the three types of experimental designs. The model result is shown in Table 4. In the model results,  $\mu_1$  represents the ratio of the variance of the error term based on the orthogonal design over the variance of the error term based on the D-efficient design.  $\mu_2$  represents the ratio of the variance of the error term based on the Bayesian design over the variance of the error term based on the D-efficient design. Similarly,  $\mu_3$  represents the ratio of the variance of the error term based on the orthogonal design over the variance of the error term based on the Bayesian method. All three values are equal to 1; no significant difference is found between the three experimental designs with respect to the variance of error term. This conclusion also supports the conclusion made by Bliemer and Rose.

**4.2. MNL Model Results Based on the Different Experimental Design Methods.** Table 5 shows the MNL model parameter estimation results for different experimental designs. According to the results of the model parameters shown in Table 5, the goodness of fit ( $\rho^2$ ) for the model based on the Bayesian D-efficient design is 0.189, which is superior to that of the D-efficient design. The goodness of fit  $\rho^2$  for the model based on the D-efficient design is 0.145, which is superior to the value of 0.14 obtained for the orthogonal design. Comparing the parameter estimation standard deviation ( $\sigma$ ) values of the

TABLE 2: The prior values of the attribute parameters.

Attributes	Parameter $\beta (T)$	Prior parameter (D-efficient)	Prior parameter (Bayesian D-efficient)
In-vehicle time	-0.036 (-4.458)	-0.036	$N (-0.036, 0.008)$
Cost	-0.038 (-2.058)	-0.038	$N (-0.038, 0.018)$
Waiting time	-0.027 (-0.361)	-0.027	$N (0, 0.076)$
Arrival time	-0.117 (-6.874)	-0.117	$N (-0.117, 0.017)$
Egress time	-0.048 (-0.604)	-0.048	$N (0, 0.080)$

TABLE 3: Statistical results of the surveyed sample population.

	Orthogonal		D-efficient		Bayesian D-efficient	
Sex						
Man: 1	162	50.6%	164	51.2%	163	50.9%
Female: 0	158	49.4%	156	48.8%	157	49.1%
Car ownership						
Yes: 1	100	31.3%	104	32.5%	102	31.8%
No: 0	220	68.7%	216	67.5%	218	68.2%
Age						
18~24	55	17.2%	56	17.5%	55	17.2%
25~50	251	78.4%	250	78.1%	252	78.7%
51~60	8	2.5%	8	2.5%	8	2.5%
>60	6	1.9%	6	1.9%	5	1.6%
Income						
<2000	54	16.9%	55	17.2%	54	16.9%
2000~4000	102	31.9%	103	32.2%	103	32.2%
4001~6000	95	29.7%	97	30.3%	96	30%
6001~8000	42	13.1%	40	12.8%	41	12.8%
>8000	27	8.4%	25	7.8%	26	8.1%
Total	<b>320</b>		<b>320</b>		<b>320</b>	

three experimental designs, it was found that the standard deviations of the model parameters based on the Bayesian D-efficient design were the smallest, followed by that of the D-efficient design. The estimated standard deviation of the model parameters based on the orthogonal design was the biggest. This suggests that both the D-efficient and Bayesian D-efficient design can achieve the theoretical goals in practice. According to the parameter estimation results, the D-error values of the three types of experimental designs were obtained, for which the orthogonal design gave a value of  $1.06 \times 10^{-4}$ , the D-efficient design was  $9.56 \times 10^{-5}$ , and the Bayesian D-efficient design was  $8.22 \times 10^{-5}$ . By comparison with the D-error value, it can be observed that the reliability of the parameter estimates based on the Bayesian D-efficient design is higher than the D-efficient design, while the D-efficient design was higher than the orthogonal design.

The D-efficient and Bayesian D-efficient design can realize the theoretical goals in the real situation, namely, minimizing the standard deviations of the estimated model parameters. To further validate the use of these two experimental designs in actual environments, this paper compared the MNL model estimated parameter standard deviations

based on the three experimental designs under different sample sizes, as shown in Figure 1.

From Figure 1, it can be observed that when using different sample sizes, the standard deviation of the estimated model parameters based on the Bayesian D-efficient design is the smallest, followed by the D-efficient design. The standard deviation of the parameters based on the orthogonal design is the biggest. These findings validate the use of the D-efficient and Bayesian D-efficient design in practice, which further verifies the conclusions made by Bliemer and Rose. The Bayesian D-efficient design can produce smaller standard deviations of parameter than D-efficient because Bayesian prior parameter is subject to certain distribution which means uncertainty by the analyst as to what the true population parameters will be empirically; the prior parameter is fixed which assumes the analyst has exact knowledge of the true population parameters when using the D-efficient design.

4.3. *ML Model Results Based on the Different Experimental Design Methods.* In this paper, the data are collected by the three stated preference questionnaires, which generated specifically for MNL model. The survey data was used to

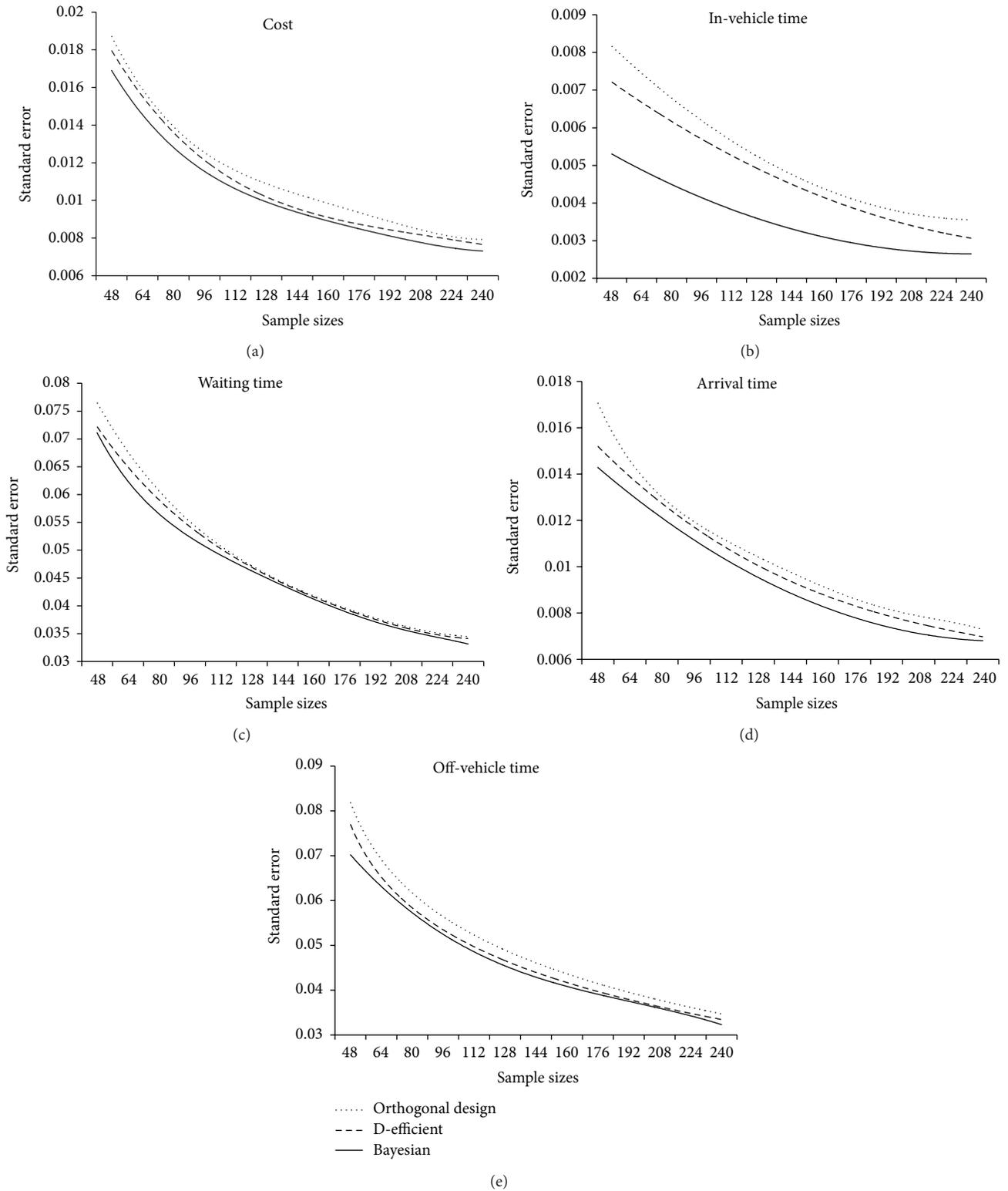


FIGURE 1: The standard deviation of the parameter estimates for different sample sizes. (a) Cost, (b) in-vehicle time, (c) waiting time, (d) arrival time, and (e) off-vehicle time.

TABLE 4: The NL model results based on the three experimental design methods.

Variables	Parameter ( $\beta_k$ )	Standard deviation ( $\sigma$ )	$T$
Cost	-0.024	0.0036	-6.685
Waiting time	-0.123	0.0164	-0.752
Arrival time	-0.095	0.0034	-27.922
In-vehicle time	-0.022	0.0014	-15.550
Off-vehicle time	-0.013	0.0165	-0.764
$\mu_1$		1.0	
$\mu_2$		1.0	
$\mu_3$		1.0	
Number of observances		3840	
LL ( $\theta$ )		-4010.982	
$\rho^2$		0.351	

TABLE 5: The MNL model results based on the different experimental design methods.

Variables	Orthogonal		D-efficient		Bayesian D-efficient	
	Parameter ( $\beta_k$ )	Standard deviation ( $\sigma$ )	Parameter ( $\beta_k$ )	Standard deviation ( $\sigma$ )	Parameter ( $\beta_k$ )	Standard deviation ( $\sigma$ )
Cost	-0.033 (-4.870)	0.0067	-0.018 (-2.882)	0.0063	-0.023 (-3.788)	0.0062
Arrival time	-0.102 (-16.012)	0.0063	-0.096 (-16.07)	0.0060	-0.096 (-16.376)	0.0058
Waiting time	-0.044 (-1.480)	0.0297	-0.003 (-0.120)	0.0297	-0.014 (-0.498)	0.0281
In-vehicle time	-0.032 (-10.580)	0.0031	-0.031 (-11.004)	0.0028	-0.013 (-5.884)	0.0022
Off-vehicle time	-0.001 (-0.031)	0.0306	-0.004 (-0.168)	0.0296	-0.018 (-0.673)	0.0280
Number of observance		1280		1280		1280
LL ( $\varphi$ )		-1354.658		-1302.555		-1294.455
$\rho^2$		0.140		0.145		0.189

estimate the ML model parameters. The ML model results are shown in Table 6.

In the ML model “cost,” “in-vehicle time,” and “arrival time” are set as the random variables in the utility functions and are subject to the  $N(\beta, \delta)$  distribution. The model results obtained based on the three experimental designs represent the travelers’ preference differences in only one variable, as shown in the ML model results given in Table 6. The symbol of estimated parameter for the attribute “off-vehicle time” in the ML model results based on the orthogonal design is wrong, but this parameter was both correct and logical in the results given by the MNL model. The ML parameter estimation symbols based on the D-efficient and Bayesian design were correct. The chi-square test between the MNL and ML model based on the two experimental designs was obtained. The chi-square test values were 10.796 and 30.253, respectively (e.g.,  $-2[LL(\varphi) - LL(\&)] = -2[-1302.555 - (-1297.157)] = 10.796$ ). Both values were greater than the chi-square distribution critical value of 7.815 under the 5%

confidence interval, indicating that the ML model results based on the D-efficient and Bayesian D-efficient design were better than MNL model. This indicates that these designs which were generated specifically for the MNL model may be not applicable to the ML model, for example, the orthogonal design. This conclusion is inconsistent with the conclusions made by Bliemer and Rose based on model simulation. They both concluded that the design generated specifically for MNL model tends to perform reasonably well when more advanced models are used. However, in this paper, it was verified based on the use of actual data that the orthogonal design may be not applicable to the complex ML model, while the D-efficient and Bayesian D-efficient design achieved accurate parameter estimation results in more complex ML model.

Based on the above analysis, the D-efficient and Bayesian D-efficient design also can achieve good parameter estimation results in the ML model. Comparing the standard deviations based on the two experimental methods in ML

TABLE 6: The ML model results based on the different experimental design methods.

Variables	Orthogonal		D-efficient		Bayesian D-efficient	
	Parameter ( $\beta_k$ )	Standard deviation ( $\sigma$ )	Parameter ( $\beta_k$ )	Standard deviation ( $\sigma$ )	Parameter ( $\beta_k$ )	Standard deviation ( $\sigma$ )
Cost						
$\beta_1$	-0.053 (-4.265)	0.0125	-0.042 (-3.111)	0.0136	-0.033 (-2.947)	0.0112
$\delta_1$	0.042 (3.440)	0.0121	0.042 (3.505)	0.012	0.025 (1.794)	0.0141
Arrival time						
$\beta_2$	-0.114 (-13.607)	0.0083	-0.104 (13.577)	0.0076	-0.229 (-6.767)	0.0339
$\delta_2$	0.005 (0.350)	0.1465	0.0011 (0.034)	0.0318	0.169 (4.845)	0.0349
Waiting time	-0.054 (-1.675)	0.0322	-0.005 (-0.166)	0.0321	-0.021 (-0.598)	0.0357
In-vehicle time						
$\beta_3$	-0.038 (-8.136)	0.0047	-0.033 (-8.536)	0.0038	-0.013 (-4.925)	0.0027
$\delta_3$	0.013 (2.337)	0.0055	0.007 (1.035)	0.007	0.004 (0.826)	0.0047
Off-vehicle time	0.008 (0.250)	0.0328	-0.008 (-0.249)	0.0321	-0.015 (-0.459)	0.0345
Number of observance	1280		1280		1280	
LL (&)	-1350.245		-1297.157		-1279.329	
$\rho^2$	0.143		0.147		0.134	

model, the conclusions drawn for the MNL model are not reflected. A degree of bias was present in some parameters, for example, the waiting time. To further investigate the validity of the theoretical advantages of these two experimental designs in different models under different sample sizes. The standard deviations of the estimated parameters between the MNL model and the ML model were compared based on the D-efficient design. As shown in Figure 2, it can be observed that the standard deviations of the ML model parameters were significantly higher than the MNL model. This shows that when using the D-efficient design, specifically for the MNL model, parameter estimation results also can be obtained accurately in the ML model; however, the theoretical advantages of the D-efficient design were not fully shown in the ML model.

*4.4. Travel Choice Behaviors Analysis from Chengdu to Longquan.* Through the above analysis, the reliability of the parameter estimation results obtained from the Bayesian D-efficient design was better. Therefore, the survey data obtained from the Bayesian D-efficient design could be chosen to calibrate the MNL model, in order to analyze the traveler behavior in the Chengdu-Longquan corridor. The model results are shown in Table 7.

The key factors affecting the traveler choice behavior in Chengdu-Longquan corridor include “cost,” “in-vehicle time,” and “arrival time.” Currently, the distance between

Chengdu and Longquan is 23 kilometers, but the average in-vehicle time for the bus is relatively long (approximately 1 hours), and the in-vehicle time for the coach is 40 minutes. Therefore, neither of these travel modes is time efficient, making “in-vehicle time” a significant factor. In the current corridor status, the number of bus lines is limited. The distribution of the bus lines is not even. Additionally, the coach is not convenient because a traveler must go to certain stations to ride the coach. However, the Chengdu subway system has not been formed into a network, so there remain significant limitations associated with riding the metro. Therefore, the variable “arrival time” became a key factor on traveler choice behavior. In the model results, the alternative constant was very significant, showing that some factors that cannot be quantified also impact the traveler choice behavior.

The personal attributes of travelers also significantly impact their choice behavior. Men exhibit a weak preference for public transport, such as the bus and metro. Compared to public transport, men prefer the use of private cars. High-income groups tend to choose private cars and taxis. In particular, private car owners tend to travel with private cars, while low-income groups are more inclined to use public transport, which has relatively low costs.

Ultimately, according to the model results, it was predicted that the rate of sharing of the five alternatives in the corridor after the opening of the metro would be 0.48% for coach, 24.47% for bus, 1.64% for taxis, 28.69% for private cars, and 44.72% for the metro. The metro is expected to take

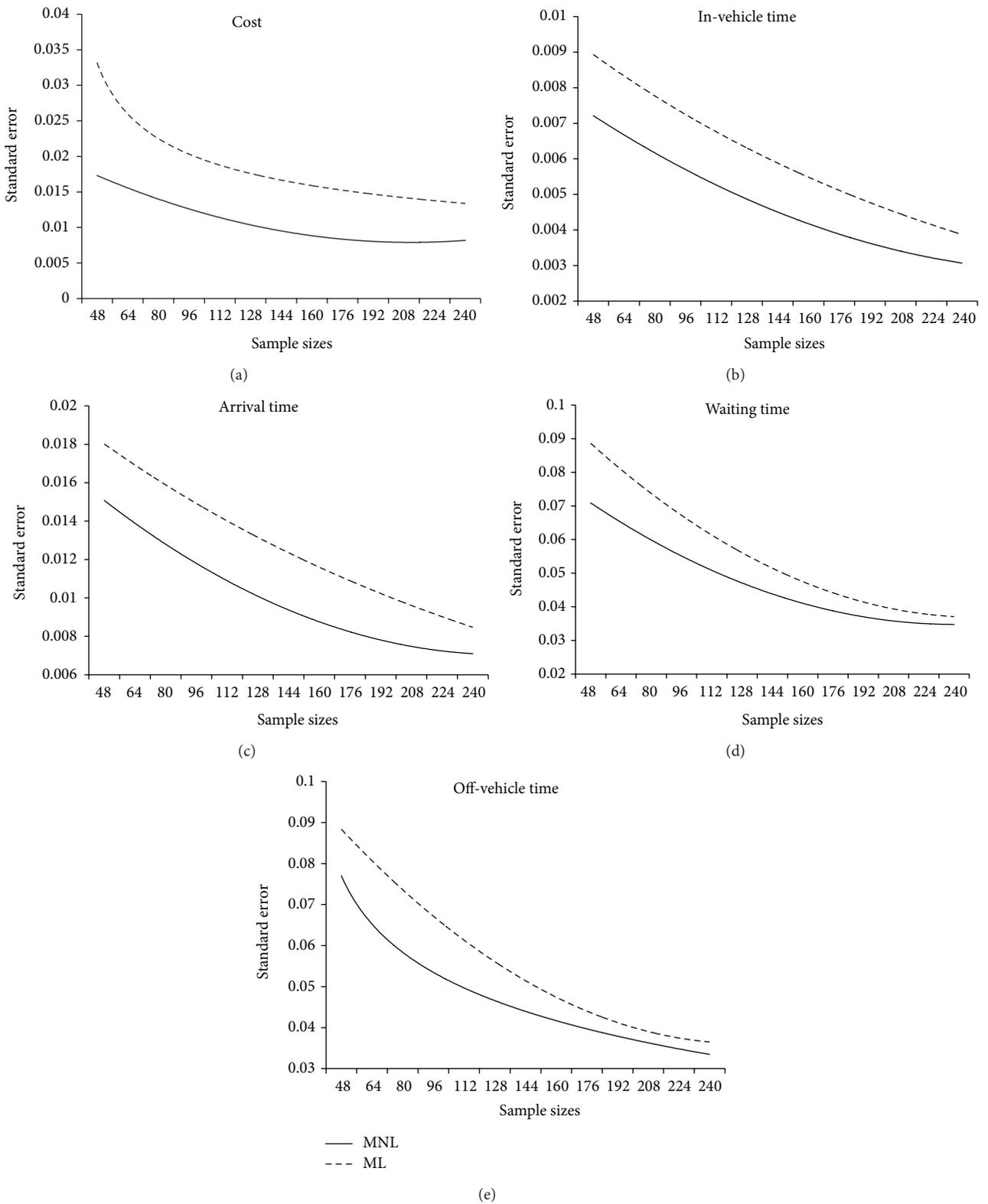


FIGURE 2: The standard deviation for the estimated parameter of the alternative attribute under different sample sizes. (a) Cost, (b) in-vehicle time, (c) waiting time, (d) arrival time, and (e) off-vehicle time.

TABLE 7: The MNL model results for analyzing the choice behavior from Chengdu to Longquan.

Variables	Parameter	<i>T</i>	Variables	Parameter	<i>T</i>
Passenger bus constant	5.163	2.734	Sex-taxi (male: 1)	-0.027	-0.047
Public bus constant	5.956	5.571	Sex-metro (male: 1)	-1.327	-5.255
Taxi constant	0.005	0.003	Age-passenger bus	-1.215	-1.182
Metro constant	7.866	7.628	Age-public bus	0.538	1.119
Cost	-0.028	-2.795	Age-taxi	-0.553	-0.751
Arrival time	-0.127	-17.106	Age-metro	-0.392	-0.856
Waiting time	-0.016	-0.486	Income-passenger bus	-0.225	-0.330
In-vehicle time	-0.024	-8.213	Income-public bus	-0.828	-4.197
Off-vehicle time	-0.018	-0.515	Income-taxi	1.138	4.371
Sex-passenger bus (Male: 1)	-1.001	-0.515	Income-metro	-0.166	-1.082
Sex-public bus (Male: 1)	-0.604	-2.130	Car ownership (Yes: 1)	6.211	14.894
LL (&)			-737.560		
Number of observances			1280		
$\rho^2$			0.481		

on a significant amount of traffic demand, demonstrating the strong appeal of metro for travelers in this corridor.

## 5. Conclusion

The experimental design is a key factor affecting the reliability of the parameters estimation in discrete choice model. There are some new experimental designs, for example, D-efficient design, but insufficient empirical research has been conducted on the effectiveness of the new designs and a little comparative analysis of the new designs against the traditional design. In the paper three kinds of stated preference questionnaires can be designed based on the three types of experimental designs, respectively. The preference data of travelers can be achieved by these stated preference questionnaires. Based on the preference data we analyze whether a difference exists in the variance of the error term for the three types of experimental designs through the application of a NL model hierarchical division. We found there is no significant difference between the three experimental designs with respect to the variance of error term. This conclusion also supports the conclusion made by Bliemer and Rose.

The MNL model parameter estimation results based on the preference data were compared. According to the results, we can demonstrate the validity of D-efficient and Bayesian D-efficient design. Bayesian prior parameter is subject to certain distribution which means the analyst is uncertain about what the true parameters will be. The prior parameter is fixed which assumes the analyst has exact knowledge of the true parameters when using the D-efficient design, but the true population parameters are not known accurately. So the Bayesian design may contain the true population parameters which performs better than the D-efficient design.

According to ML model results, these designs which were generated specifically for the MNL model may be not

applicable to the advanced ML model, for example, the orthogonal design. This conclusion is inconsistent with the conclusions made by Rose and Bliemer based on model simulation. Compared to the parameter estimation results of the MNL model based on the D-efficient design with the ML model under different sample sizes, we find the theoretical advantages of the D-efficient design were not fully shown in the ML model. But in this paper, we only used ML model; the other advanced model is not considered.

The preference data obtained from the Bayesian D-efficient design can be chosen to estimate the MNL model parameters. The traveler choice behavior in Chengdu-Longquan corridor was analyzed. These attributes, such as “cost,” “arrival time,” “in-vehicle time,” and “income,” can impact significantly the traveler choice behavior. These results reflect that there are some shortcomings of bus on convenience and timeliness. Finally, we predict the metro can handle over 44% passenger flow if the metro will be operated in the future.

## Conflict of Interests

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

## Acknowledgments

This work was supported by National Science Foundation of China under Grant nos. 50908195 and 51178403, Specialized Research Fund for the Doctoral Program of Higher Education (no. 20130184110020) and the Fundamental Research Funds for the Central Universities (no. SWJTU11CX080 and no. 2682014CX130), Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University (no. K201207), Program for New Century Excellent Talents in University (NCET-13-0977), The National Basic Research

Program of China (973 Program no. 2012CB725405), and the Science and Technology Innovation Practice Program for Graduate Student, Southwest Jiaotong University (no. YC201407119), and Science & Technology Department of Sichuan Province (no. 2014RZ0037).

## References

- [1] Z. Sándor and M. Wedel, "Designing conjoint choice experiments using managers' prior beliefs," *Journal of Marketing Research*, vol. 38, no. 4, pp. 430–444, 2001.
- [2] V. V. Can, "Estimation of travel mode choice for domestic tourists to Nha Trang using the multinomial probit model," *Transportation Research A: Policy and Practice*, vol. 49, pp. 149–159, 2013.
- [3] A. Hoen and M. J. Koetse, "A choice experiment on alternative fuel vehicle preferences of private car owners in the Netherlands," *Transportation Research A*, vol. 61, pp. 199–215, 2014.
- [4] D. Hoyos, "The state of the art of environmental valuation with discrete choice experiments," *Ecological Economics*, vol. 69, no. 8, pp. 1595–1603, 2010.
- [5] D. A. Hensher, J. M. Rose, and W. H. Greene, *Applied Choice Analysis: A Primer*, Cambridge University Press, Cambridge, UK, 2005.
- [6] T. Fowkes, "The design and interpretation of freight stated preference experiments seeking to elicit behavioural valuations of journey attributes," *Transportation Research B: Methodological*, vol. 41, no. 9, pp. 966–980, 2007.
- [7] J. M. Rose and M. C. Bliemer, "Stated preference experimental design strategies," in *Handbook of Transport Modelling*, pp. 151–180, Elsevier, Oxford, UK, 2008.
- [8] B. J. Kanninen, "Optimal design for multinomial choice experiments," *Journal of Marketing Research*, vol. 39, no. 2, pp. 214–227, 2002.
- [9] M. C. J. Bliemer, J. M. Rose, and D. A. Hensher, "Efficient stated choice experiments for estimating nested logit models," *Transportation Research B: Methodological*, vol. 43, no. 1, pp. 19–35, 2009.
- [10] M. C. J. Bliemer and J. M. Rose, "Experimental design influences on stated choice outputs: an empirical study in air travel choice," *Transportation Research A: Policy and Practice*, vol. 45, no. 1, pp. 63–79, 2011.
- [11] J. J. Louviere, T. Islam, N. Wasi, D. Street, and L. Burgess, "Designing discrete choice experiments: do optimal designs come at a price?" *Journal of Consumer Research*, vol. 35, no. 2, pp. 360–375, 2008.
- [12] A. Tudela and G. Rebolledo, "Optimal design of stated preference experiments when using mixed logit models," in *Proceedings of the European Transport Conference (ETC' 06)*, Leiden, The Netherlands, 2006.
- [13] D. A. Hensher and J. M. Rose, "Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: a case study," *Transportation Research A: Policy and Practice*, vol. 41, no. 5, pp. 428–443, 2007.
- [14] D. A. Hensher, J. M. Rose, J. D. D. Ortúzar, and L. I. Rizzi, "Estimating the willingness to pay and value of risk reduction for car occupants in the road environment," *Transportation Research A: Policy and Practice*, vol. 43, no. 7, pp. 692–707, 2009.
- [15] M. C. J. Bliemer and J. M. Rose, "Construction of experimental designs for mixed logit models allowing for correlation across choice observations," *Transportation Research B: Methodological*, vol. 44, no. 6, pp. 720–734, 2010.
- [16] M. Bliemer and J. M. Rose, "Designing stated choice experiments: state-of-the-art," in *Proceedings of the 11th International Conference on Travel Behaviour Research*, Kyoto, Japan, 2006.
- [17] R. Kessels, B. Jones, P. Goos, and M. Vandebroek, "Recommendations on the use of Bayesian optimal designs for choice experiments," *Quality and Reliability Engineering International*, vol. 24, no. 6, pp. 737–744, 2008.



# Hindawi

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

