Journal of Choice Modelling 49 (2023) 100436

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Research note

Sample size selection for discrete choice experiments using design features

Samson Yaekob Assele [a,*], Michel Meulders [b], Martina Vandebroek [a]

[a] *Faculty of Economics and Business, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium*
[b] *Faculty of Economics and Business, KU Leuven, Warmoesberg 26, 1000 Brussels, Belgium*

ARTICLE INFO

Keywords:
Sample size calculation
Discrete choice experiment
D-optimal design
Rule of thumb
Linear and quantile regression

ABSTRACT

In discrete choice experiment (DCE) studies, selecting the appropriate sample size remains a challenge. The question of the required sample size for a DCE is addressed in the literature in two distinct approaches: a rule-of-thumb approach and an approach based on the statistical error of the parameter of interest. The former is less accurate and does not depend on the desired power and significance level, whereas the latter requires knowing the complete design which may not be known at the planning stage. This paper proposes a new rule of thumb as well as a new regression-based method that requires knowing certain design characteristics rather than the complete design and takes into account the power and significance level. We compare the sample size estimated using the proposed methods with the true required sample size based on the statistical error of the parameter of interest and the approximations given by the existing rules of thumb. The results show that both the new rule of thumb and the regression-based approach improve the magnitude and proportion of underestimation compared to the most commonly used rule of thumb of Orme. Though the proposed approaches perform in general similarly to Tang's rule which improves Orme's rule, they seem to do better for large settings in terms of the number of choice sets and the number of alternatives per choice set in reducing underestimation. Moreover, we have demonstrated the possibility to adapt the regression-based approaches to take into account other scenarios and choice set complexity.

1. Introduction

In Discrete Choice Experiments (DCE), as in any other statistical study, the precision of the parameter estimates and the power for hypothesis testing are influenced by the size of the sample. If a large sample can easily be obtained, the statistical power and precision may not be of much concern. However, given limited resources for recruiting survey respondents, the sample sizes used in practice are often limited by the available budget. Thus, it is crucial to know whether the sample size used is adequate to retrieve statistically significant parameter estimates with enough power.

There are different approaches to calculate the required sample size for a DCE. These include rules of thumb (Pearmain and Gleave, 1991; Orme, 1998; Johnson and Orme, 2003; Hensher et al., 2005; Tang et al., 2006; Lancsar and Louviere, 2008), exact power calculations based on either the statistical error of the choice probabilities (Louviere et al., 2000) or the statistical error of the preference parameters (Rose and Bliemer, 2013; de Bekker-Grob et al., 2015) and an approximate approach based on regression models using design features and design efficiency (Yang et al., 2015). Section 2 provides a brief review of these approaches.

* Corresponding author.
E-mail address: samsonyaekob.assele@kuleuven.be (S.Y. Assele).

https://doi.org/10.1016/j.jocm.2023.100436
Received 20 October 2022; Received in revised form 6 June 2023; Accepted 27 July 2023
Available online 8 September 2023
1755-5345/

An important aspect of calculating the required sample size of a DCE, is the role that the experimental design plays. For a given sample size, the ability to retrieve statistically significant parameter estimates depends on the experimental design chosen by the researcher. For instance, Rose and Bliemer (2013) shown that D-efficient and S-efficient designs outperform traditional orthogonal designs in terms of obtaining significant parameter estimates, provided that reliable prior information is available to generate the design. Additionally, Kessels et al. (2006) compared D-, A-, G-, and V-optimality criteria for constructing Bayesian designs. Their findings indicate that the distinct optimality criteria do not significantly differ in estimation performance, but the V-optimal designs and, the G-optimal designs are indeed best suited for predictive purposes.

In this paper, we propose a new rule of thumb and a new regression-based approach for estimating the required sample size assuming that a D-optimal design will be used. In these approaches, the sample size calculation is based on certain design features rather than on the entire design. The DCE design features used include the number of choice sets, the number of alternatives per choice set, the number of parameters, and the maximum number of levels for dummy-coded attributes. In this regard, Rose and Bliemer (2013) have also demonstrated the influence of various design features on the sample size requirements of stated choice experiments. They showed that the number of attribute levels and the attribute level range have a significant impact on the accuracy of the parameter estimates. They also showed that increasing the number of choice tasks improves statistical efficiency (in terms of decreasing D-error or S-error). However, for a multinomial logit (MNL) model, the marginal gain in overall efficiency from using larger designs is limited once the choice tasks that provide the most information are included in the design. On the other hand, for a panel mixed logit (MIXL) model, using a larger design contributes relatively more to the efficient estimation of the heterogeneity parameters. Similar results are also reported by Vanniyasingam et al. (2018) in their systematic review paper on how the DCE design characteristics are related to the statistical efficiency based on 9 eligible studies. Furthermore, Kessels et al. (2006) also illustrated that a three-alternative optimal Bayesian design yields more accurate parameter estimates, compared to a two-alternative design.

There have been several attempts to determine the required sample size for DCE studies using one or more of the design features. The earliest attempt by Orme (1998) provided a formula to approximate the sample size needed based on the number of choice tasks, the number of alternatives, and the largest number of levels for any of the attributes. This formula was revised by Tang et al. (2006) to better represent the heterogeneity of the sample and the complexity of the model. Recently, Yang et al. (2015) have used regression models to calculate the sample size for DCE health studies using several design features and design efficiency. Although the rules of thumb are not accurate, they have been commonly used in several fields. For example, a review by de Bekker-Grob et al. (2015) shows that of 69 healthcare DCE studies that were conducted since 2012, 13% used one or more of the rules of thumb proposed by Orme (1998), by Pearmain and Gleave (1991), or by Lancsar and Louviere (2008) to determine the sample size.

The purpose of this paper is twofold. First, we propose a new rule of thumb and a regression-based approach that uses design features to predict the required sample size in order to obtain significant estimates for all the parameters in the DCE. Second, we examine the reliability of the proposed methods and the existing rules of thumb. The rest of the paper is organized as follows. In Section 2 we review the different existing approaches to conduct sample size calculations for DCEs. In Section 3, we propose two new ways to calculate the sample size: a new rule of thumb and a regression-based approach. In Section 4, we describe the results of the regression-based approach and compare its performance with the new and existing rules of thumb. We end with a discussion and conclusion in Section 5.

2. Literature review

Sample size calculations for Discrete Choice Experiment (DCE) studies have been approached in different ways. Researchers often used rules of thumb based on past experience (Pearmain and Gleave, 1991; Orme, 1998; Johnson and Orme, 2003; Hensher et al., 2005; Lancsar and Louviere, 2008). The most commonly used rule of thumb to determine the required sample size for a DCE was proposed by Orme (1998) (see also Johnson and Orme (2003)). The rule indicates that the number of respondents N should satisfy the inequality

$$N \geq 500\frac{L^{max}}{JS}, \tag{1}$$

where L^{max} is the largest number of levels for any of the attributes, J is the number of alternatives per choice task (not including the "None"), and S is the number of choice sets each respondent receives. When considering all two-way interactions in the utility, L^{max} is equal to the largest product of the number of levels of any two attributes. Tang et al. (2006) provided a revised version of Orme's formula. Rather than the constant "500" in Orme's formula, which was interpreted as a measure of sample variability, a variable index I is used to capture the heterogeneity of the sample. Besides, instead of using L^{max}, which (Tang et al., 2006) interpreted as a measure of model complexity, they used the number of parameters to be estimated. The revised heuristic is given as

$$N \geq I\frac{K}{\tilde{J}S(1-c)}, \tag{2}$$

where I is the Homogeneity Index, c is the expected percentage of choice sets where respondents choose the none alternative, K denotes the number of parameters to be estimated, and $\tilde{J}$ is a recoding of the number of alternatives per choice set, i.e. $\tilde{J} = \{2, 2.5, 3, 4\}$ for $J = \{2, 3, 4, 5+\}$. Tang et al. (2006) recommended an I value of 500 for general population studies and a smaller value of 100 for specialized and therefore more homogeneous populations.

There are other recommendations for the required sample size based on past experience with similar studies. For instance, Pearmain and Gleave (1991) suggested a sample size of at least 100 respondents to obtain reliable estimates for preference data whereas (Orme, 2010) tripled the minimum number of respondents to 300 for robust quantitative research based on common practices in the market research community. On the other hand, Hensher et al. (2005) recommended that a minimum of 50 respondents should choose each alternative in a labeled design. Lancsar and Louviere (2008) mentioned the use of 20 respondents per version of blocked designs to reliably estimate models and indicated the need for a larger sample size to estimate covariate effects in a post hoc analysis. Note that these rules of thumb do not account for the desired precision or power level and are assumed to work well for any purpose.

Louviere et al. (2000) proposed a parametric approach to determine the minimum acceptable sample size for a DCE based on the sampling error for proportions. This approach assumes a priori knowledge of the choice proportions associated with each alternative and is therefore not applicable for determining the minimum required sample size for stated choice experiments with unlabeled alternatives.

More recent papers address the issue of minimum sample size requirements in terms of the precision of the parameter estimates (Rose and Bliemer, 2013; de Bekker-Grob et al., 2015). These approaches require the derivation of the Asymptotic Variance-Covariance matrix (AVC), which depends on the statistical model, the prior parameter values, and the design. Details on how to construct the AVC for the MNL are given in McFadden (1974) and for the panel MIXL in Bliemer and Rose (2010).

Given the AVC matrix, Rose and Bliemer (2013) look for the minimum sample size for which all parameters are statistically significant. Following the notation in Rose and Bliemer (2013), let β_k, $\hat{\beta}_k$ and $\bar{\beta}_k$ denote the true parameter value for attribute k, its estimate and the prior value, respectively. The theoretical lower bound of the sample size for finding a statistically significant parameter estimate for β_k assuming a deviation in one direction from zero is

$$N_k > \left[\frac{Z_{1-a} \cdot se(\hat{\beta}_k)}{\hat{\beta}_k}\right]^2, \tag{3}$$

where a is the significance level[1], and Z_{1-a} is the $100(1-a)$th quantile of the standard normal distribution. In this approach, estimates for $se(\hat{\beta}_k)$ are approximated as the square root of the kth diagonal element of the AVC based on the prior parameter values, resulting in asymptotic standard errors $se(\bar{\beta}_k)$. So the minimum sample size is computed by

$$N_k > \left[\frac{Z_{1-a} \cdot se(\bar{\beta}_k)}{\bar{\beta}_k}\right]^2. \tag{4}$$

The sample size calculation of Rose and Bliemer (2013) is focused on establishing confidence intervals for parameter estimates that exclude zero. In a more general approach, de Bekker-Grob et al. (2015) look for the minimum sample size for testing the significance of the parameter with a desired power level. The minimum required sample size to reject the hypothesis: $H_0 : \beta_k = 0$ vs. $H_a : \beta_k > 0$ or $H_0 : \beta_k = 0$ vs. $H_a : \beta_k < 0$ at the significance level a and the power level $1 - b$ can be calculated as

$$N_k > \left[\frac{(Z_{1-a} + Z_{1-b}) se(\bar{\beta}_k)}{\bar{\beta}_k}\right]^2, \tag{5}$$

where Z_{1-a} and Z_{1-b} represent the $100(1-a)$th and $100(1-b)$th quantiles of the standard normal distribution, respectively. Remark that the approach of Rose and Bliemer (2013) can be seen as a special case of the approach of de Bekker-Grob et al. (2015) with $b = 50\%$. Although it is possible to determine the required sample size based on any $\bar{\beta}_k$, here we specify the true required sample size based on the most critical parameter where $N^{true} = \max_k(N_k)$.

Yang et al. (2015) developed a formula to determine the sample size in terms of the required precision of the estimated utility difference between the profiles with the largest and smallest utility, the design characteristics and design efficiency. The formula is derived from a regression model that associates the utility difference precision obtained from re-sampling various sample sizes from 32 actual DCE studies (34 data sets) with the corresponding drawn sample size, design characteristics and experimental design efficiency. Yang et al. (2015) demonstrate that, unlike the other approaches discussed above, their method accounts for both statistical and measurement errors by using the actual DCE data. Moreover, their approach focuses on estimating the required sample size for utility difference precision rather than the preference parameter precision used in the exact approaches (Rose and Bliemer, 2013; de Bekker-Grob et al., 2015). On the other hand, the requirement to specify the desired utility difference precision and experimental design efficiency on top of the design characteristics makes this approach more complicated as compared to the other rules of thumb, which only depend on design characteristics. As with the other rules of thumb, this approach also does not depend on the desired significance and power level. It, however, provides a new perspective for determining sample size taking into account both statistical and measurement errors based on regression models fitted on re-sampled cases from actual data. In Section 4.4, we discuss how to take into account choice set complexity when determining the required sample size for a desired precision of preference parameters.

The practical applicability of the different approaches discussed so far depends on which information is available. The traditional rules of thumb require information on only some aspects of the design, but they are generally not very accurate. In these approaches, two designs will yield the same required sample size if the design features included in the rule of thumb are equal, even though the designs could differ in other design features. The exact approaches discussed by Rose and Bliemer (2013) and de Bekker-Grob et al. (2015) on the other hand involve computation of the AVC, which requires knowledge of the model, the prior parameter values, and

[1] If the direction of the deviation does not matter i.e. if negative and positive values of the estimated parameter are considered theoretically possible, $a/2$ should be used instead of a in formulas (3), (4) and (5).

the DCE design. Part of this information is usually unknown at the planning stage. In the absence of complete design information, the rule-of-thumb approach can be seen as a rough approximation to determine the minimum sample size required. In this paper, we aim to develop approaches that only rely on design features and that better approximate the true minimum sample size than the existing rules of thumb.

3. Methods

3.1. A new rule of thumb

As an alternative to the existing rules of thumb, we propose and evaluate a very simple rule based on the total number of parameters (K) and the number of choice tasks per respondent (S). In this rule, we specify the number of respondents in such a way that for each parameter of the choice model to be estimated there should be at least T choice tasks, namely:

$$N \geq T\frac{K}{S}. \tag{6}$$

The value of the constant T is determined from the simulated data discussed in the subsequent section. This rule can be seen as a revision of Tang et al. (2006) when the number of alternatives J is unknown. We refer to this rule hereafter as the T choices per parameter (TCPP) rule as NS gives the total number of choices made by N respondents in S choice sets and K is the total number of parameters.

3.2. A regression-based approach

Inspired by the approach of Yang et al. (2015), we also use a regression model to approximate the ratio $\frac{se(\bar{\beta}_k)}{\bar{\beta}_k}$ in Eq. (5) based on design features, where $\bar{\beta}_k$ is the prior value of the critical parameter (i.e. the parameter most difficult to estimate),

$$\frac{se(\bar{\beta}_k)}{\bar{\beta}_k} = \mathbf{x}'\boldsymbol{\alpha} + \epsilon, \tag{7}$$

where $\mathbf{x}$ is a vector of design characteristics, $\boldsymbol{\alpha}$ is the vector of parameters, and ϵ is the error term. We then compute the minimum sample size N by predicting the value of $\frac{se(\bar{\beta}_k)}{\bar{\beta}_k}$ using the fitted model and plugging this predicted value in Eq. (5).

This approach ensures that on average the required sample size will be correctly estimated but also that in approximately 50% of the cases the study will be underpowered. Therefore, we also use quantile regression (QR) models. If we choose a high quantile (eg. 80% quantile) we can reduce the underestimation. In analogy with the traditional linear regression models, the linear quantile regression model introduced by Koenker and Bassett (1978) specifies a linear dependence of conditional quantiles of y on x. The QR model with the τth quantile for the response takes the form of

$$\frac{se(\bar{\beta}_k)}{\bar{\beta}_k} = \mathbf{x}'\boldsymbol{\alpha}^{(\tau)} + \epsilon^{(\tau)} \tag{8}$$

where the superscript τ, $0 < \tau < 1$, denotes a quantile of the response, $\boldsymbol{\alpha}^{(\tau)}$ is a vector of quantile specific parameters, and $\epsilon^{(\tau)}$ is the quantile specific error term. The regression coefficient vector $\boldsymbol{\alpha}^{(\tau)}$ is estimated by minimizing

$$\sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i\boldsymbol{\alpha}^{(\tau)}), \tag{9}$$

where $\rho_\tau(.)$ is the loss function defined by $\rho_\tau(u) = u(\tau - I(u < 0))$ and $I(.)$ denotes the indicator function. A comprehensive description of the theory, application, and interpretation of quantile regression can be found in Koenker (2005). Furthermore, the quantile regression model can be estimated using, for instance, the R package "quantreg" (Koenker, 2022).

The design characteristics that we consider in the linear and quantile regression models include the number of choice tasks, the number of alternatives per choice set, the maximum number of attribute levels for any of the dummy-coded attributes (a value of 0 is used if the attribute is not categorical), and the total number of parameters.

To fit the linear and quantile regression models, data is simulated. Given that the experimental design chosen by the analyst plays a significant role in getting accurate estimates, we assume that the researcher will use a D-optimal design assuming non-zero prior parameter values. We generate the designs using the R-package idefix (Traets et al., 2020) for various design characteristics. We selected the values of design characteristics based on an overview of 25 studies conducted by Akinc Abdulhayoglu (2019). The overview included studies that were conducted after the year 2000, that collected real-life data using DCE with effect-coded attribute levels, that estimated a MIXL model, and that are among the most cited articles from several different fields, such as Transportation, Health Economics, Consumers Studies etc. Appendix A presents a summary of this overview and Fig. A.1 in Appendix A shows the absolute value of the estimated preference parameters of the studies in the overview. The range of the levels used for each design characteristic in the simulation is given in Table 1.

The simulation includes the following steps:

- Randomly select a combination of the number of attributes (P), the number of alternatives (J), and the number of choice sets (S).
- For the selected combination, randomly specify the type of each attribute as discrete or continuous.

Table 1
Design characteristics used in simulation.

Design characteristics	Range
Number of attributes (P)	4–8
Number of attribute levels	2–5
Number of alternatives (J)	2–6
Number of choice sets (S)	6, 8, 9, 10, 12, 14, 16

Table 2
Estimates of the OLS and 70th quantile regression models.

Coefficients	OLS			QR ($\tau = 0.7$)		
	Estimate	Std. error	P-value	Estimate	Std. error	P-value
Intercept	2.202	0.130	0.000	2.131	0.202	0.000
Nb of alternatives	−0.129	0.018	0.000	−0.096	0.028	0.001
Nb of choice sets	−0.112	0.011	0.000	−0.126	0.017	0.000
Max nb of levels of dummy-coded attrts.	0.426	0.035	0.000	0.437	0.055	0.000
Nb of parameters	0.175	0.022	0.000	0.250	0.034	0.000
R-squared:	0.09					

- If the selected attribute is discrete, the number of levels is determined by randomly selecting from 2 to 4 levels. If the attribute is continuous, the number of levels is selected from 3 to 5, and the value of the levels are coded in an equally spaced fashion from −1 to 1.
- Check whether $S \geq K + 1$, where K is the number of parameters. If not, return to step 1 as the model cannot be estimated.
- Select non-zero prior parameter values from a uniform distribution $U(-2, -0.1)$ or $U(0.1, 2)$.
- Generate the local D-optimal design for the main effects model using idefix.
- Compute the AVC using the model, the prior parameter values, and the design.
- Take $\max_k \left(se(\bar{\beta}_k)/\bar{\beta}_k\right)$.

This maximum of $se(\bar{\beta}_k)/\bar{\beta}_k$ across all values of k is used as the dependent variable of the regression model as this is the most critical parameter requiring the largest sample size.

We generated 5916[2] D-optimal designs by varying design characteristics and prior parameter values. Designs that result in a sample size of less than 10 respondents at 5% significance level and 90% power were excluded. Note that the recommended sample sizes in DCE using the rules of thumb range also from a minimum of 20 respondents per questionnaire version (Lancsar and Louviere, 2008) to a minimum of 300 respondents (Orme, 2010). We are left with 5199 simulated designs for further analysis. Appendix B provides an overview of the characteristics of all simulated designs. We can see from this Appendix that there seems to be no major difference in design characteristics between the excluded and included designs except that the excluded designs have a relatively large value of the critical parameter, which explains the resulting small sample size of less than 10 in the excluded designs.

The simulations were used to study the relationship between the design characteristics and the ratio of the asymptotic standard error to the prior parameter value $se(\bar{\beta}_k)/\bar{\beta}_k$. We estimated several OLS and quantile regression models considering interactions as well as transformations of the dependent variable and of the regressors. Including the interaction terms and using the transformations yielded only a slight improvement in model fit as well as in predictive performance, hence we use the untransformed linear main-effects models for the sake of simplification.

The coefficients of the OLS and of 70th quantile regression model are given in Table 2, whereas the Table in Appendix C presents estimates for 50th and 60th quantile regression models. It is clear from the OLS estimates in Table 2 that all the design characteristics included in the model have a significant impact. An increase in the number of parameters and the maximum number of attribute levels for dummy-coded attributes increases the ratio $se(\bar{\beta}_k)/\bar{\beta}_k$, which is directly proportional to the estimated sample size. In contrast, the number of alternatives per choice set and the number of choice sets are negatively related to $se(\bar{\beta}_k)/\bar{\beta}_k$ as an increase in these design features decreases the estimates of the required sample size. The estimates from the 50th, 60th, and 70th quantile regression models have equally intuitive parameters except that there is a change in the magnitude of the coefficients.

The estimated models are then used to predict $se(\bar{\beta}_k)/\bar{\beta}_k$. Both the true and the predicted $se(\bar{\beta}_k)/\bar{\beta}_k$ are used to compute the required sample sizes for a specific significance level a and power $1 - b$ as described in Section 3, resulting in the truly required and the predicted sample size.

[2] We set the simulation to 8000 designs, but limited the maximum time to compute the D-optimal design to 10 min yielding 5916 designs.

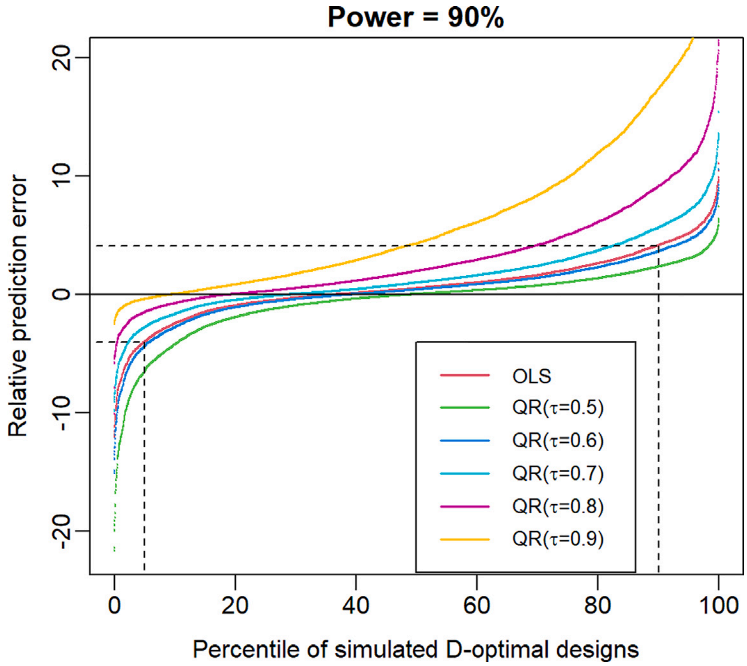


Fig. 1. RPE of sample sizes computed based on the linear and quantile regression models using a 5% significance level and 90% power.

4. Results

4.1. Relative prediction error

To evaluate the performance of the existing and the new approaches, we calculated the relative prediction error as follows:

$$RPE = \frac{N^{pred} - N^{true}}{min(N^{pred}, N^{true})}, \tag{10}$$

where RPE stands for relative prediction error, N^{true} is the true required sample size given the specified power and significance level assuming $\bar{\beta}_k = \beta_k$, and N^{pred} is the sample size computed using either the existing rules of thumb or based on the proposed methods. Taking the minimum of the predicted and the true sample size in the denominator of Eq. (10) provides a similar scale for visual comparison of underestimation and overestimation. The results however need to be interpreted carefully: for instance, an RPE of 2 indicates an overestimation of twice the true value, i.e. the predicted value is three $(= RPE + 1)$ times the true value, whereas an RPE equal to -2 indicates an underestimation of twice the predicted value, i.e. the true value is three $(= |RPE| + 1)$ times the predicted value.

4.2. Performance of the regression-based approach

We first compute and compare the relative prediction error (RPE) of the sample sizes obtained by the regression approach. In this approach, we fitted the OLS and the 50th, 60th, 70th, 80th and 90th quantile regression models. Fig. 1 depicts the RPE for the sample sizes determined based on the regression models at a 5% significance level and 90% power. Remark that for other choices of the significance and power level the relative position of the RPE remain exactly the same as can be seen from Fig. 2.

The plot in Fig. 1 is similar to the fraction of design space (FDS) plots used in response surface design (Zahran et al., 2003). It shows the relative prediction error on the y-axis versus the percentiles of the simulated D-optimal designs having a relative prediction error below the y-value of the curve on the x-axis. We can see for instance that the predicted sample size based on the OLS model can become higher than 5 $(RPE = 4)$ (see the dashed lines) times the true sample size for 10% of the simulated designs when there is overestimation and the true required sample size needs more than 5 $(RPE = -4)$ times the predicted sample size for about 5% of the simulated designs in case of underestimation. We also see that for the OLS results, approximately 35% of the true sample sizes are underestimated as indicated by the point where the OLS curve crosses the x-axis. Underestimation of the true sample size is relatively high for the OLS model and for the 50th, and 60th quantile regression models compared to the models based on the other quantiles. The proportion and the magnitude of underestimation of the true sample size in the simulated data drop when higher quantile regression models are used whereas the magnitude of the overestimation increases with the higher quantile regression models. It seems that sample size selection by the 70th quantile regression model slightly reduces the proportion of underestimation while only moderately increasing the overestimation.

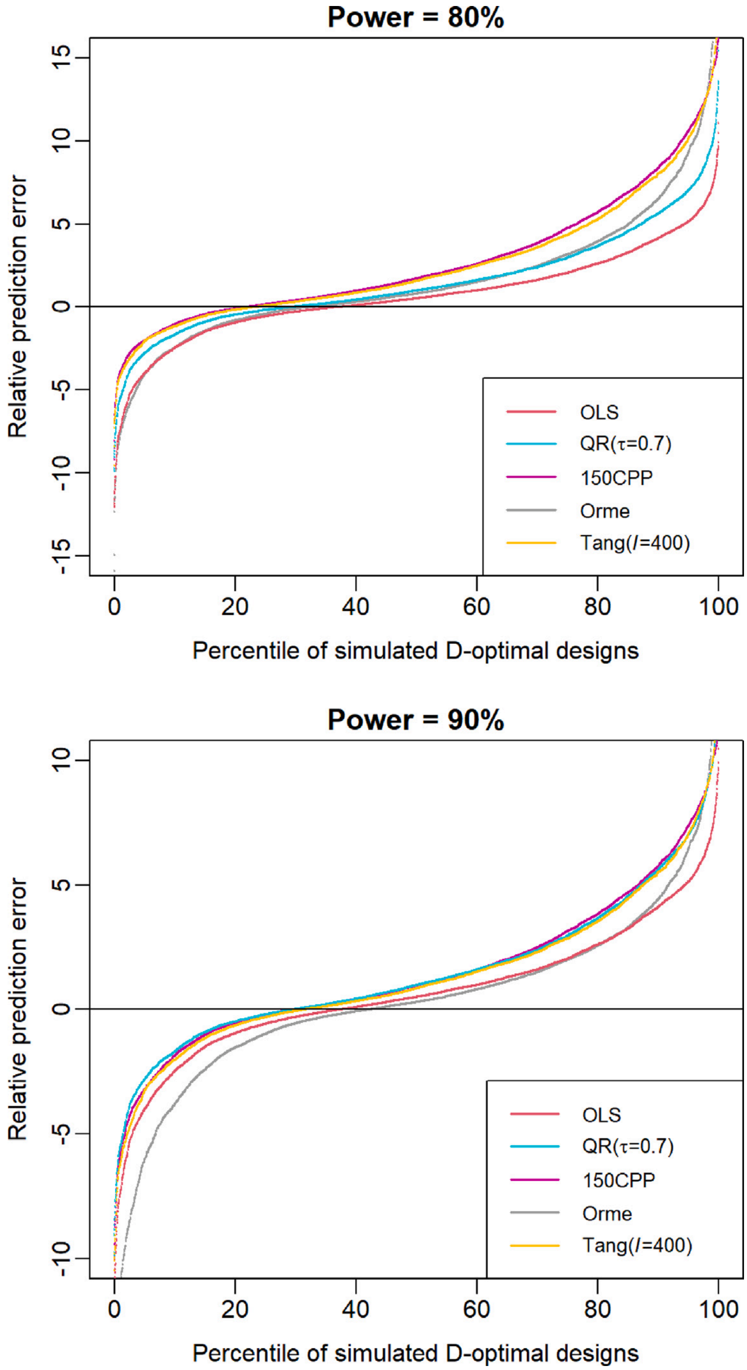


Fig. 2. RPE of sample sizes computed using OLS, $QR(\tau = 0.7)$, the 150CPP rule, and the existing rules of thumb by Orme and by Tang (I = 400). The true and the predicted sample sizes by the OLS, and the $QR(\tau = 0.7)$ methods are computed using a 5% significance level with a power of 80% in the top panel and with a power of 90% on the bottom panel.

4.3. Comparison of the regression-based approach with the new and existing rules of thumb

We also compare the performance of our approaches with the existing rules of thumb by Orme (1998) and its generalization by Tang et al. (2006). We computed the required sample size for each design using the different approaches discussed. For the regression-based approach (OLS and $QR(\tau = 0.7)$), the sample size was computed using a 5% significance level with 80% and 90% power. For our proposed rule of thumb and the existing rule of thumb by Tang et al. (2006), the sample size was computed for

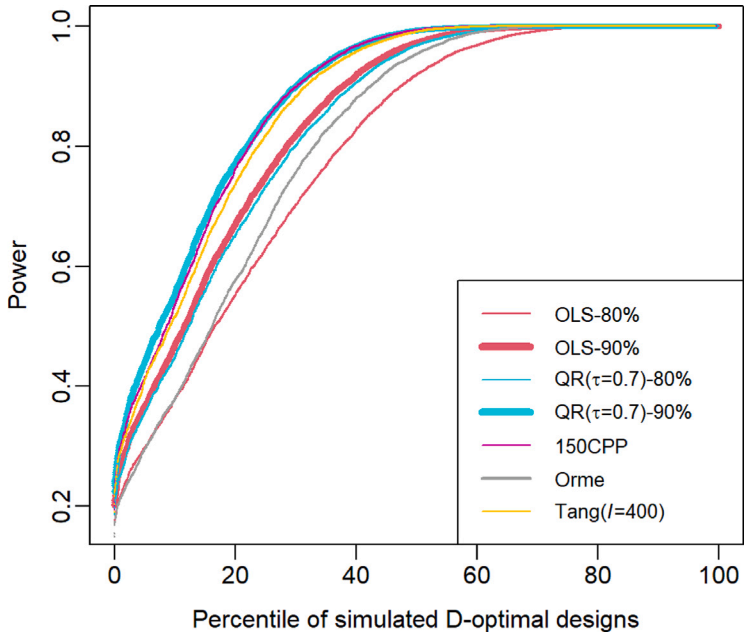


Fig. 3. The power of the designs with the sample sizes computed using OLS and $QR(\tau = 0.7)$ models, the 150CPP rule, and the existing rules of thumb by Orme and by Tang (I = 400). Sample sizes are calculated for the OLS and $QR(\tau = 0.7)$ models at a 5% significance level with 80% and 90% power.

different values of the constants T and the index I. The RPE of these sample sizes was calculated relative to the true sample size required for 80% and 90% power with the 5% significance level.

For our comparison in this section, we choose the constant $T = 150$ for our new rule of thumb after comparing the RPE (at 90% power) for different values of the constant T. Fig. D.1 in Appendix D provides this comparison. We refer to this rule hereafter as the 150 choices per parameter (150CPP) rule. Similarly, for Tang's approach, we used 400 as the value for the index I. As can be seen from Fig. D.2 in Appendix D, the recommended values for the index I (i.e 100 or 500) do not seem appropriate for our simulation setting (resulting in 75% underestimation or an overestimation of more than 5 times the true sample size for about 25% of the simulated designs).

It is clear from the top panel of Fig. 2 that for a 5% significance level and 80% power, the proportion of underestimation is similar in the $QR(\tau = 0.7)$ and the rule of thumb by Orme. The magnitude of underestimation and overestimation however is slightly lower for $QR(\tau = 0.7)$ as can be seen from the bottom left and the top right corner of the figure. At this power level, Tang's approach and the 150CPP rule has the lowest proportion and magnitude of underestimation, but they have the highest overestimation. Overall, $QR(\tau = 0.7)$ seems to provide the best balance. The bottom panel in Fig. 2 on the other hand shows that when we increase the required power level to 90% while keeping the 5% significance level, the $QR(\tau = 0.7)$, Tang ($I = 400$) and the 150CPP seem to have a similar prediction performance on average. These three approaches reduce the proportion of underestimated sample sizes by 10% compared to Orme, and by 5% compared to OLS. It is interesting to see that the existing simple rules of thumb which do not depend on the power level perform somewhat similar to the regression approaches.

More insights can be found from Fig. 3 which depicts the power of the sample sizes computed using the different approaches discussed using the true values of $se(\bar{\beta})$ and $\bar{\beta}$. On the y-axis are the power values and on the x-axis are the percentiles of the simulated D-optimal designs with power below the y-value of the curve. The proportion of the simulated designs having less than 80% power ranges from 20 to 40 percent. The figure also shows the benefit of taking the desired power into account in the regression-based approach. The wide gap between the curve of $QR(\tau = 0.7) - 90\%$ and the curve of $Orme$ in this figure demonstrates the advantage of our approach when the desired power is high. There is however only a slight advantage in using the $QR(\tau = 0.7)$ approach compared to the 150CPP and Tang's approach which generalizes Orme's rule of thumb.

Despite the similarity in overall performance of these approaches, the sample size computed for a given design can vary. Appendix E includes the sample sizes computed using the different approaches for various ranges of true N (at 90% power) based on all included simulated designs. Furthermore, for these designs, Fig. E.1 in Appendix E also shows box plots with the distribution of the true sample size and the distribution of the calculated sample size for each method. We can see from this Appendix that, while all approaches estimate very modest sample sizes, which can approximate the true required sample size for most designs, the actual required sample size can be very large for a smaller set of designs (15%). Besides, as there is a significant portion of underestimation of the true sample size in all these approaches (see Fig. 2), we further investigate the scenarios in which the regression approaches or the other rules of thumb which depend on the design characteristics are giving useful results.

Fig. F.1 in Appendix F shows that for larger D-optimal designs in terms of the number of choice sets ($S \geq 12$) or the number of alternatives per choice set ($J \geq 4$), the $QR(\tau = 0.7)$ method performs better in reducing the underestimation of the true sample size compared to the other approaches. Similarly, the $QR(\tau = 0.7)$ method performs better for designs with a large number of parameters ($K > 6$) or when the maximum number of levels for dummy-coded attributes is large ($L_d^{max} \geq 4$). For smaller D-optimal designs, the benefit of the $QR(\tau = 0.7)$ approach in reducing the underestimation is not that large compared to the other approaches.

As the prior parameter value for the critical parameter has a significant role in determining the required sample size, it is interesting to assess how the considered approaches perform in terms of the value of the critical parameter. Fig. G.1 in Appendix G presents the relative prediction errors (at 90% power) for sample sizes calculated using the proposed and existing rules of thumb for three ranges of the critical value where there is over-prediction, modest over/underprediction, and severe underprediction. It also shows the relative prediction error that was computed by using the smallest sample size, the largest sample size, and the average sample size across all of the methods. It is clear that when the critical $\bar{\beta}_k$ is larger than 0.3, the different regression models considered hardly underestimate the true minimum sample size. In this case, using Orme's rule of thumb or the regression-based OLS approach or taking the minimum predicted sample size of all the approaches can reduce the overestimation. For critical $\bar{\beta}_k$ between 0.2 and 0.3, the $QR(\tau = 0.7)$ method seems appropriate having the lowest underestimation and an average overestimation compared to the rest. Finally, for critical parameter values smaller than 0.2, there is severe underprediction in all the approaches. Although at this level the regression-based approach $QR(\tau = 0.7)$ performs better than the rest, at least 3 times ($RPE = -2$) the estimated sample size is needed to correctly approximate the true sample size for about 20% of the designs with critical parameters less than 0.2. Thus for small critical values, the rule-of-thumb approaches must be used with caution.

4.4. Accounting for choice set complexity

By re-sampling from data sets of several actual studies, Yang et al. (2015) captured several sources of error variance occurring in practice while estimating the required sample size for utility difference precision. Similarly, as the precision of preference parameters can also be negatively affected by increased error variance due to increases in choice set complexity or other sources of error, it is important to take into account these errors while estimating the required sample size. Although it is challenging to take into consideration all factors that influence the precision of the estimates at the planning phase of the sample size, it is feasible to account for potential measurement errors related to the choice task complexity. For instance, Danthurebandara et al. (2011) discuss how to take the effect of choice set complexity into account when constructing choice experiments whereas (DeShazo and Fermo, 2002) illustrate how to take into account choice set complexity when analyzing the data using the heteroscedastic conditional logit (HCL) model. Using the HCL model, it is also possible to account for choice set complexity when determining the true sample size required. Note that we do not assume that data have been collected yet, but that we can take the impact of choice set complexity into account.

In estimating the truly required sample size using the exact approaches in this paper, we have used the AVC matrix determined as the inverse of Fisher information matrix for a MNL model, as discussed earlier. For the MNL model, the Fisher information matrix I^{MNL} for N = 1 is given by

$$I^{MNL}(\beta|\mathbf{X}) = \sum_{s=1}^{S} \mathbf{X}_s'(\mathbf{P}_s - \mathbf{p}_s\mathbf{p}_s')\mathbf{X}_s, \tag{11}$$

where $\mathbf{X}_s$ is the design matrix for choice set s, $\mathbf{P}_s$ is a diagonal matrix with diagonal elements $\mathbf{p}_s$, $\mathbf{p}_s = (p_{1s}, \dots, p_{Js})$ is a vector of choice probabilities for each alternative in choice set s, and p_{js} is the probability that an alternative j is chosen in a choice set s,

$$p_{js} = \frac{exp(\mu \mathbf{x}_{js}' \boldsymbol{\beta})}{\sum_{l=1}^{J} exp(\mu \mathbf{x}_{ls}' \boldsymbol{\beta})}, \tag{12}$$

where μ is the scale factor usually assumed to be constant, $\mathbf{x}_{js}$ is a K-dimensional vector containing the attribute values of alternative j in choice set s, $\boldsymbol{\beta}$ is a K-dimensional vector of parameters.

For the HCL model, the choice probability is given by

$$p_{js}^{HCL} = \frac{exp(\mu(C_s)\mathbf{x}_{js}' \boldsymbol{\beta})}{\sum_{l=1}^{J} exp(\mu(C_s)\mathbf{x}_{ls}' \boldsymbol{\beta})}, \tag{13}$$

where $\mu(C_s)$ is the scale as a function of C_s which measures the complexity of the choice set s. Although there exist other complexity measures to take complexity into account, here we used entropy introduced by Swait and Adamowicz (2001) as a measure of choice set complexity. To capture the U-shaped relation between the scale parameter and choice set complexity (see Swait and Adamowicz (2001) and Danthurebandara et al. (2011)), we use the following exponentiated quadratic function of entropy H_s,

$$\mu(C_s) = exp(-0.25H_s + 0.5H_s^2). \tag{14}$$

The entropy is defined as

$$H_s = -\sum_{s=1}^{S} p_{js} ln(p_{js}) \tag{15}$$

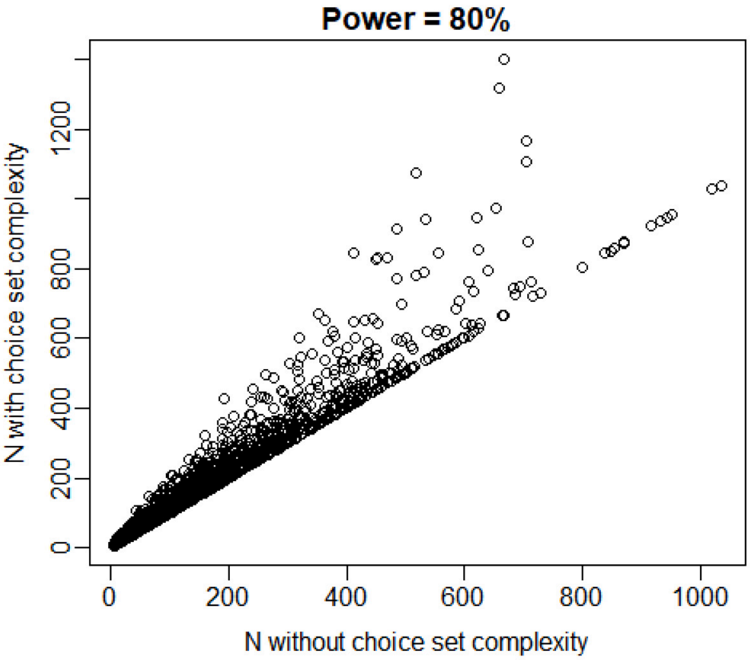


Fig. 4. Scatter plot of the true required sample size computed with and without taking choice set complexity into account.

with p_{js} as defined in Eq. (12).

The details for computing the Fisher information matrix of the HCL model can be found in Danthurebandara et al. (2011) and Sándor and Franses (2009). For illustration purposes, we used the Fisher information for the MNL given in Eq. (11), with the HCL probability in Eq. (13) to derive the AVC, which in turn is used to compute the true required sample size taking into account the choice set complexity.

Fig. 4 presents the true sample size calculated with and without taking design complexity into account. As expected, we need larger sample sizes in the presence of choice set complexity to maintain the same power level. As an illustration, Fig. 5 shows the RPE of sample sizes estimated using the $QR(\tau = 0.7)$ model when the true sample size is calculated with and without taking the choice set complexity into account. As expected, the underestimation gets larger as the true sample sizes became larger.

5. Discussion and conclusion

In a DCE, a reasonable specification of the sample size is necessary as it affects the accuracy of the model estimates and the cost of the study. Although the exact methods discussed by Rose and Bliemer (2013) and de Bekker-Grob et al. (2015) provide accurate estimates of the required sample size provided that an accurate estimate of the AVC is available, usually little is known at the planning stage of the study which necessitates alternative approaches for sample size specification. Unfortunately, there is no simple alternative that can give an accurate estimation of the required sample size. Ideally, the sample size should not be underestimated to avoid type II errors and overestimation should be kept to a minimum to reduce the cost of data collection.

In this paper, we explored new methods to calculate the sample size using DCE design features. These include an alternative rule of thumb based on the number of parameters and the number of choice tasks per respondent and a regression-based approach. In the regression-based approach, we use linear and quantile regression models to predict the ratio $se(\bar{\beta}_k)/\bar{\beta}_k$ of the critical parameter based on the DCE design characteristics, which in turn is used to determine the required sample size. Finally, we examined the performance of the proposed approaches and the existing rules of thumb in approximating the true sample size needed.

The regression approaches provide an additional option to calculate the required sample size for DCE. For the simulation settings considered in this study, the approaches based on the OLS and $QR(\tau = 0.7)$ are able to reduce the proportion of underestimated

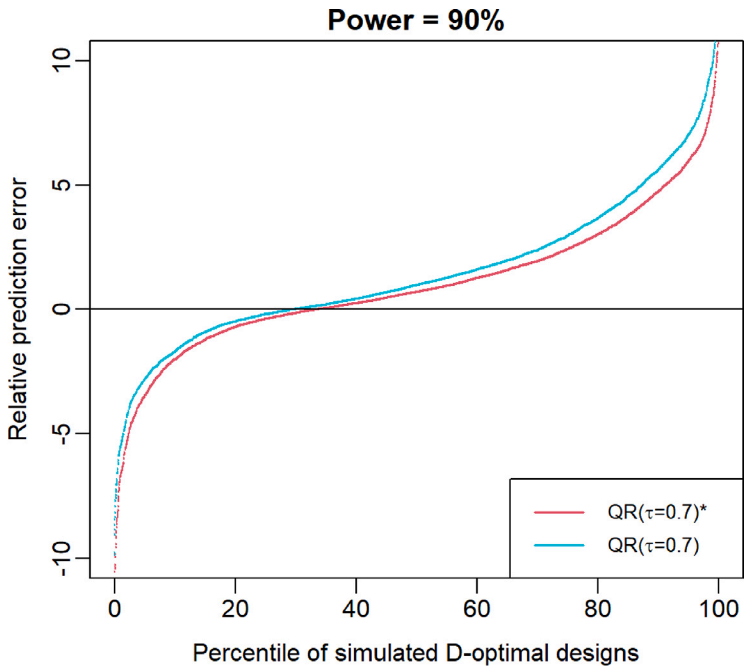


Fig. 5. RPE of sample sizes computed using $QR(\tau = 0.7)$. The * indicates that the RPE is computed relative to the true sample size taking choice set complexity into account.

sample sizes (with a slight increase in the magnitude of overestimation) compared to the rule of thumb by Orme. Similarly, the 150CPP rule and the rule of Tang(I = 400) reduce the proportion and magnitude of underestimation compared to the rule of thumb by Orme. Clearly, in any of these approaches, reducing underestimation of the required sample size for some cases results in more overestimation for other cases. Moreover, the proportion of underestimated designs that ranges from 20 to 40 percent in both the proposed and the existing approaches is an indication that such rules of thumb need to be used with caution.

The regression-based approach can easily be adapted to different settings. For instance, if there is prior information on preference parameter values with a narrower range than considered here, the coefficients of the OLS and the quantile regression model can be obtained for that range which can give a more accurate approximation compared to the existing rules of thumb. It is also possible to tune these coefficients by taking choice set complexity into account. This can easily be done by re-fitting the regression models discussed on the response ($se(\bar{\beta}_k)/\bar{\beta}_k$) derived from the AVC matrix for the HCL model that takes the choice set complexity into account. The R-code for conducting the regression-based approach can be obtained from the authors upon request.

In summary, the accuracy of the selected sample size depends on the information available at the time of calculating the sample size. When only limited information about the design is available, the rules of thumb can provide a rough approximation to the required sample size. The alternative approaches presented in this paper can supplement the existing rules of thumb and can take the significance level and the required power into account. Using different approaches helps to get an idea of the range of the required sample size. For design settings with a relatively large choice set size or a large number of alternatives per choice set or a large number of parameters the $QR(\tau = 0.7)$ method performs better in reducing the underestimation. The results also demonstrate that information regarding the size of the critical parameter is crucial in determining the usefulness of the proposed and existing rules of thumb as we found only adequate sample sizes if $\bar{\beta}_k \geq 0.3$ for attributes that are scaled between -1 and 1.

Further research can be conducted based on this study. The current results are useful when the sample size has to be selected to obtain significant estimates for all the parameters of a MNL model which is estimated using a local D-optimal design, assuming correct specification of the prior values. Testing the applicability of the results and/or methods for other problem settings is left for future research. One topic to be investigated is the criterion to be used as the study objective may also include the estimation of willingness to pay, or the estimation of utility differences, or of other transformations based on the preference parameter estimates. Based on the required precision, this will result in different required sample sizes. Furthermore, prior misspecification while generating local D-optimal designs or the use of orthogonal or random designs will increase the required sample size but the precise impact still has to be investigated. Last but not least, it is very important to note that the study was restricted to the MNL model whereas the mixed logit model, latent class model, nested logit model, or other extensions of the MNL model will generally require a larger sample. In conclusion, new sample size selection methods were proposed and contrasted with existing methods. The prediction errors, which were presented in FDS plots for easier comparison, show once again that it remains challenging to predict the required sample size without detailed information on the parameters and design.

CRediT authorship contribution statement

Samson Yaekob Assele: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Michel Meulders:** Conceptualization, Methodology, Supervision, Writing – reviewing and editing. **Martina Vandebroek:** Conceptualization, Methodology, Supervision, Writing – reviewing and editing.

Declaration of competing interest

None

Funding

This research was funded by the C14/17/016 - DBOF project of KU Leuven and the G0C7317N - FWO project

Appendix A. Summary of design characteristics from reviewed studies

Article	Research field	Citations	N	S	J	P	L^{max}
1	Consumer Studies	114	200	8	4	5	
2	Food Economics	25	159	8	3	4	3
3	Agricultural Research	54	155	9	3	5	3
4	Food Economics	74	261	32	2	8	3
5	Tourism and Cultural Economics	174	785	5	4	8	
6	Environment Economics	13	325	16	4	4	5
7	Energy Policy	43	400	4	3	5	
8	Health Economics	86	400	4	3	4	
9	Health Economics	116	210 from Jewish 261 from general	16	4	12	4
10	Health Research	59	200	10	2	3	3
11	Health Economics	175	409	8	2	6	2
12	Transportation	16	251	9	3	4	4
13	Ecological Economics	96	473	12	3	5	5
14	Health Policy	41	511	8	2	5	2
15	Ecological Economics	18	736 772 763	12	2	3	3
16	Health Research	17	1117	5	3	4	3
17	Transportation	6	93	6	3	7	4
18	Health Research	13	161 males 325 females	9	2	5	4
19	Agricultural Economics	134	253	12	3	8	3
20	Health Research	11	206	9	3	4	3
21	Food Economics	91	145 for beef 150 for chicken	12	6	8	
22	Health Policy	8	594	6	2	5	
23	Agriculture	9	500 399	8	3	4	3
24	Consumer Studies	23	1489	12	3	3	4
25	Ecological Economics	22	200	6	3	6	5

Where N is the sample size used, S is the choice set size, J is the number of alternatives per choice set, P is the number of attributes, L^{max} is the maximum number of levels for any of the attributes.

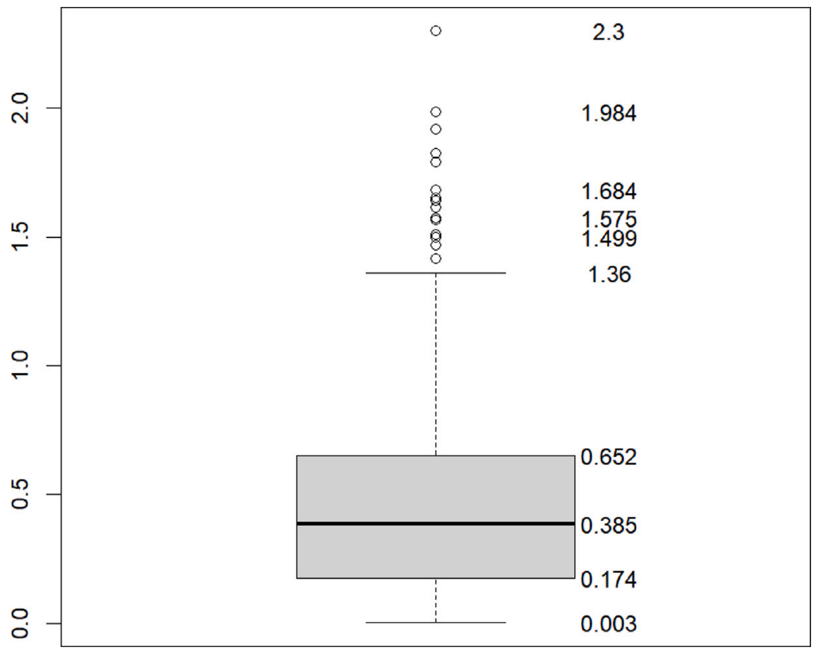
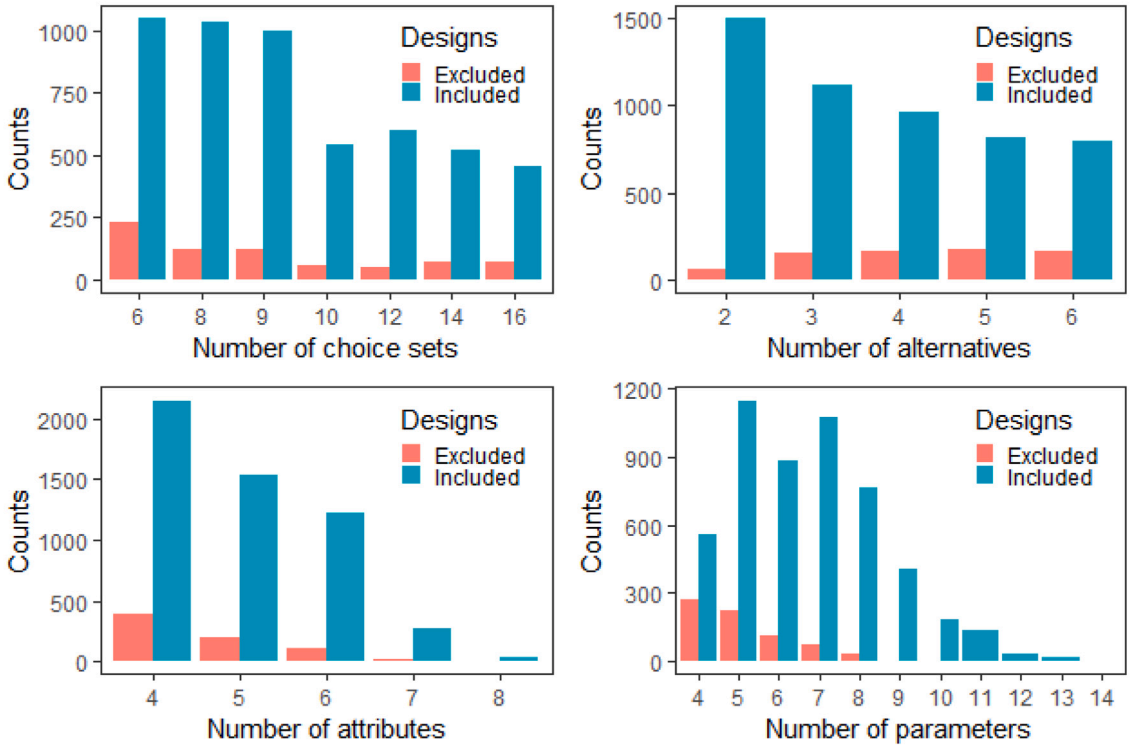
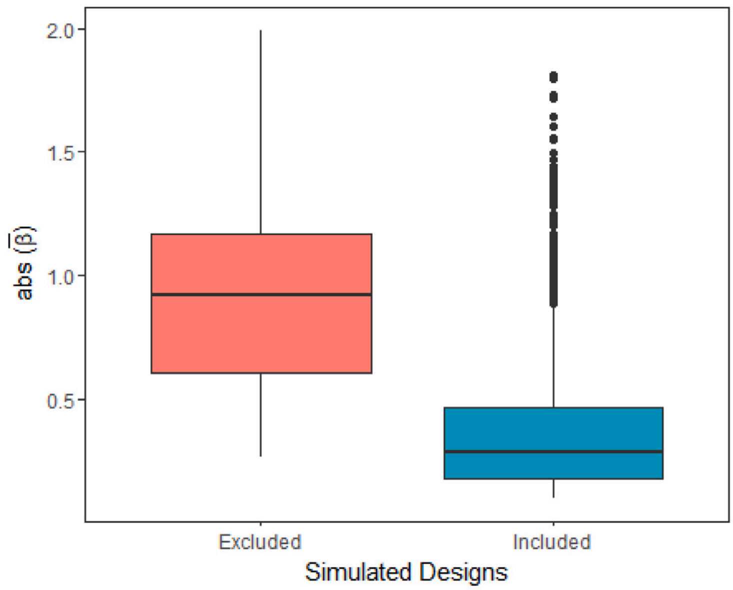


Fig. A.1. Absolute value of estimates for preference parameters based on the 25 studies reviewed.

Appendix B. Characteristics of simulated designs





Appendix C. Coefficients for 50th and 60th quantile regression models

Coefficients	QR ($\tau = 0.5$)			QR ($\tau = 0.6$)		
	Estimate	Std. error	P-value	Estimate	Std. error	P-value
Intercept	1.504	0.126	0.000	1.771	0.163	0.000
Nb of alternatives	−0.077	0.018	0.000	−0.079	0.023	0.001
Nb of choice sets	−0.073	0.011	0.000	−0.095	0.014	0.000
Max nb of levels of dummy-coded attrts.	0.241	0.034	0.000	0.315	0.044	0.000
Nb of parameters	0.191	0.021	0.000	0.218	0.028	0.000

Appendix D. Sensitivity analysis of TCPP and tang's rule of thumb

See Figs. D.1 and D.2.

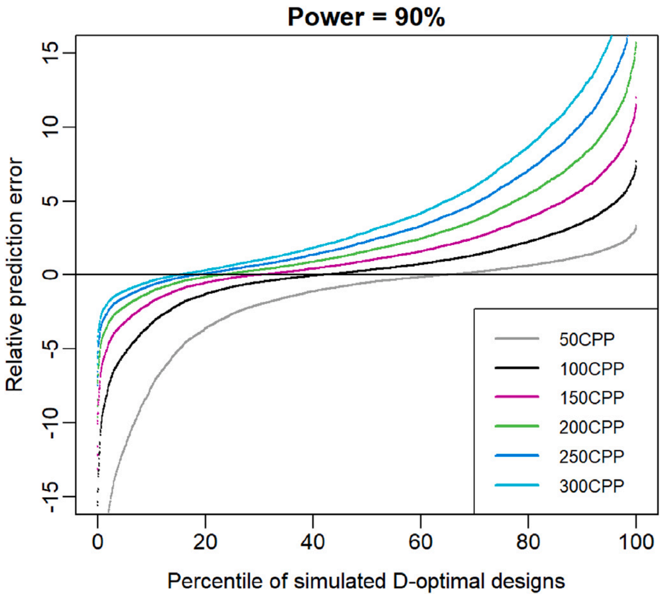


Fig. D.1. RPE of sample sizes computed based on the new rule of thumb (TCPP) for different values of the constant T relative to the true sample size computed using a 5% significance level and 90% power.

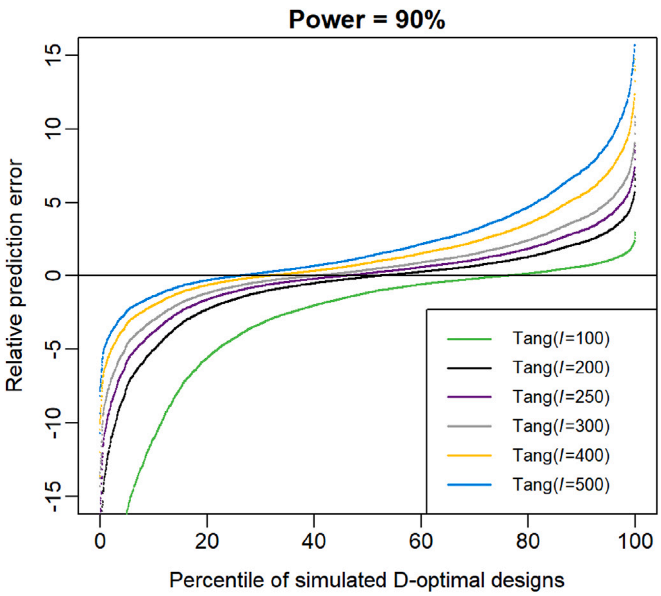


Fig. D.2. RPE of sample sizes computed based on Tang's rule of thumb for different values of the index I relative to the true sample size computed using a 5% significance level and 90% power.

Appendix E. Sample sizes computed based on different approaches for various ranges of true N (90% power)

See Fig. E.1.

N (Ranges)	Freq of designs	OLS	QR ($\tau = 0.7$)	Orme	Tang	150CPP
		5th, 50th, 95th	5th, 50th, 95th	5th, 50th, 95th	5th, 50th, 95th	5th, 50th, 95th
(10, 30]	1621	26, 74, 120	36, 97, 160	29, 69, 167	50, 89, 160	60, 100, 131
(30, 50]	893	33, 86, 132	43, 112, 170	31, 69, 167	50, 100, 167	64, 105, 133
(50, 100]	1067	36, 88, 135	45, 114, 177	31, 69, 156	50, 100, 175	66, 107, 133
(100, 150]	473	48, 93, 140	62, 120, 182	31, 69, 156	56, 107, 175	67, 113, 133
(150, 200]	275	48, 96, 135	62, 123, 179	29, 71, 208	63, 111, 168	75, 113, 133
(200, 300]	341	51, 100, 140	66, 129, 186	28, 63, 156	50, 107, 175	67, 113, 135
(300, 400]	198	60, 100, 140	76, 129, 187	33, 78, 156	62, 112, 178	75, 113, 135
(400, 500]	110	53, 108, 140	68, 140, 189	36, 78, 139	64, 120, 178	83, 117, 135
(500, 1430]	221	72, 105, 144	93, 136, 187	42, 83, 167	75, 120, 178	84, 120, 135

Where 5th, 50th, 95th are the percentile values.

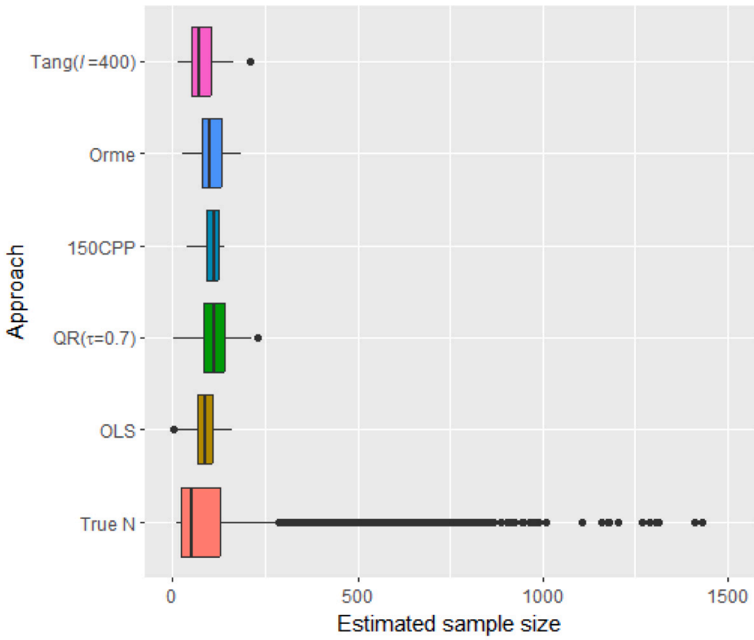


Fig. E.1. Distribution of True N and distribution of calculated sample sizes for each method based on all included simulated designs.

Appendix F. RPE by number of alternatives (J) and by number of choice sets (S)

See Fig. F.1.

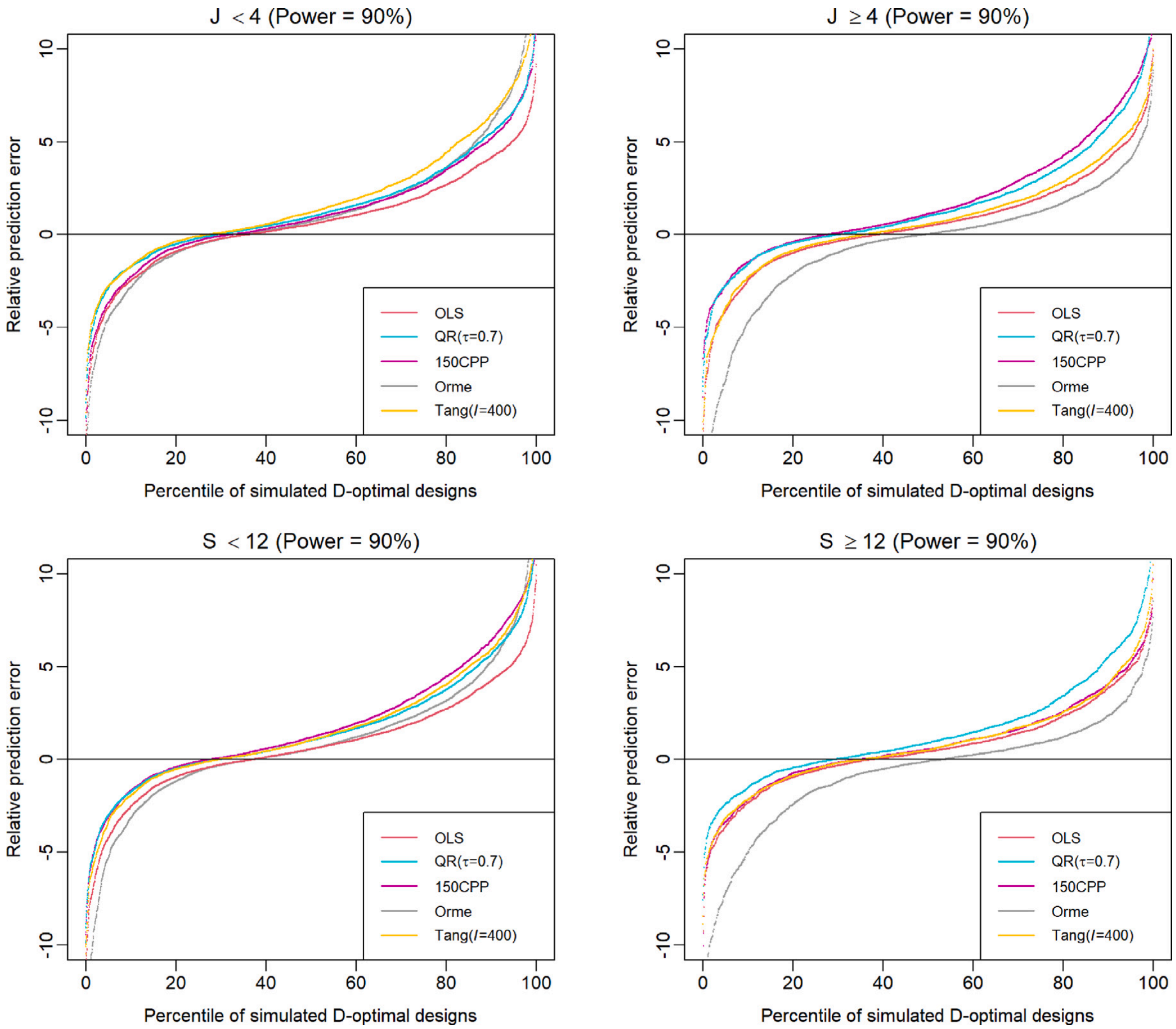


Fig. F.1. RPE of sample sizes computed using OLS, $QR(\tau = 0.7)$, the 150CPP rule, and the existing rules of thumb by Orme and by Tang (I = 400) by the number of alternatives J (top panel) and the number of choices set S (bottom panel). The true and the predicted sample sizes by the OLS, and the $QR(\tau = 0.7)$ methods are computed using a 5% significance level with a power of 90%.

Appendix G. RPE by the value of critical parameter

See Fig. G.1.

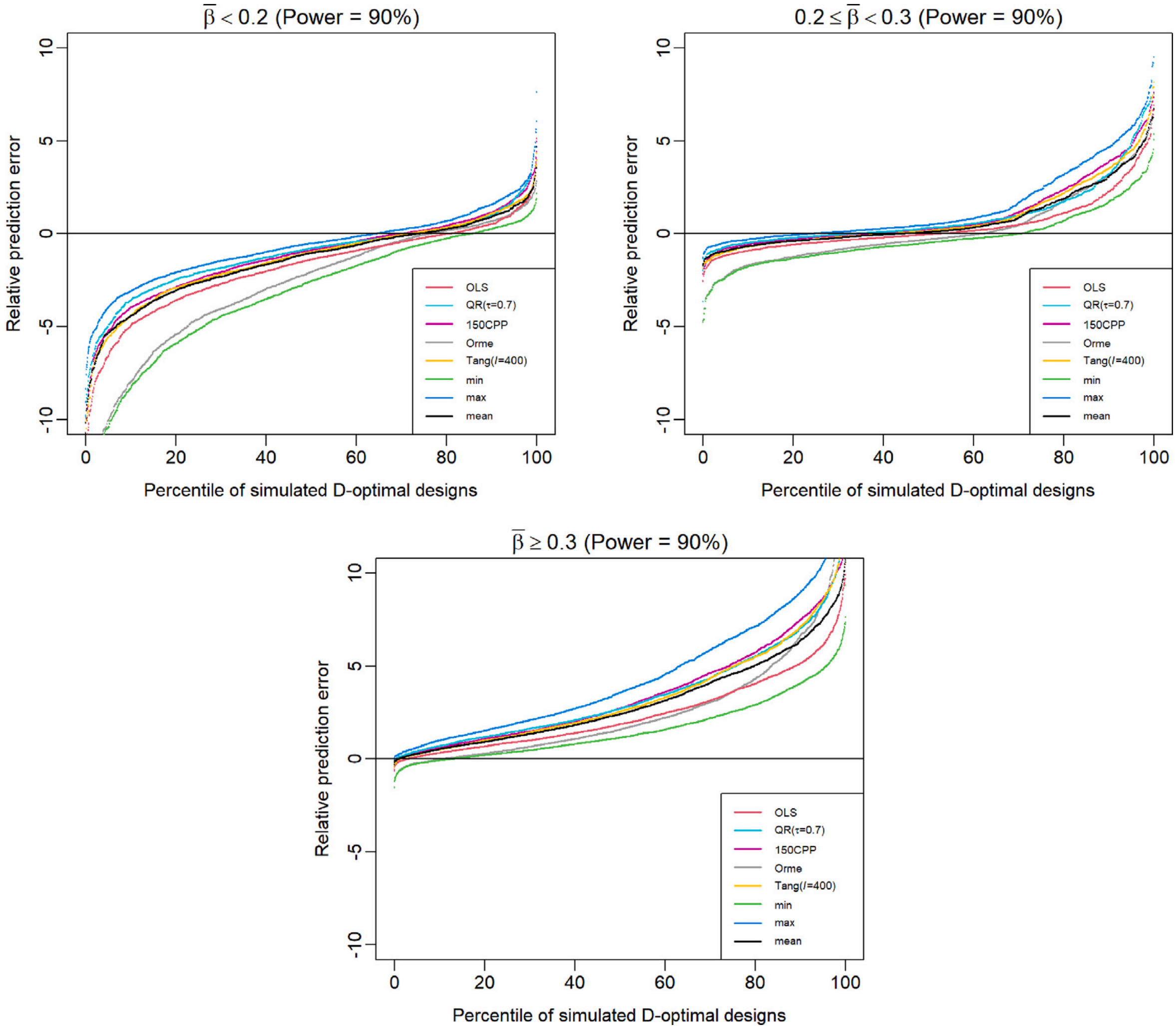


Fig. G.1. RPE of sample sizes computed using OLS, $QR(\tau = 0.7)$, the 150CPP rule, and the existing rules of thumb by Orme and by Tang (I = 400) by ranges of critical parameter value $\bar{\beta}$. The true and the predicted sample sizes by the OLS, and the $QR(\tau = 0.7)$ methods are computed using a 5% significance level with a power of 90%.

References

Akinc Abdulhayoglu, D., 2019. Contributions to the Design and the Analysis of Choice Experiments (Ph.D. thesis). KU Leuven.

Bliemer, M.C., Rose, J.M., 2010. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Transp. Res. B 44 (6), 720–734.

Danthurebandara, V.M., Yu, J., Vandebroek, M., 2011. Effect of choice complexity on design efficiency in conjoint choice experiments. J. Statist. Plann. Inference 141 (7), 2276–2286.

de Bekker-Grob, E.W., Donkers, B., Jonker, M.F., Stolk, E.A., 2015. Sample size requirements for discrete-choice experiments in healthcare: A practical guide. Patient - Patient-Cent. Outcomes Res. 8, 373—384. http://dx.doi.org/10.1007/s40271-015-0118-z.

DeShazo, J., Fermo, G., 2002. Designing choice sets for stated preference methods: The effects of complexity on choice consistency. J. Environ. Econ. Manage. 44 (1), 123–143.

Hensher, D.A., Rose, J.M., Rose, J.M., Greene, W.H., 2005. Applied Choice Analysis: A Primer. Cambridge university press, Cambridge.

Johnson, R., Orme, B., 2003. Getting the Most from CBC. Sawtooth Software research Paper Series 98382.

Kessels, R., Goos, P., Vandebroek, M., 2006. A comparison of criteria to design efficient choice experiments. J. Mar. Res. 43 (3), 409–419. http://dx.doi.org/10.1509/jmkr.43.3.409.

Koenker, R., 2005. Quantile Regression. In: Econometric Society Monographs, Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511754098.

Koenker, R., 2022. Quantreg: Quantile regression [computer software manual]. Retrieved from https://CRAN.R-project.org/package=quantreg, (R package version 5.88).
Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46 (1), 33–50.
Lancsar, E., Louviere, J., 2008. Conducting discrete choice experiments to inform healthcare decision making: A user's guide. PharmacoEconomics 26, http://dx.doi.org/10.2165/00019053-200826080-00004.
Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. Stated Choice Methods: Analysis and Applications. Cambridge university press, Cambridge.
McFadden, D.L., 1974. In: Zarembka (Ed.), Conditional Logit Analysis of Qualitative Choice Behavior. In: Frontiers in Econometrics, pp. 105–142.
Orme, B., 1998. Sawtooth Software Sample Size Issues for Conjoint Analysis Studies. Research Paper Series 98382.
Orme, B., 2010. Sawtooth Software Sample Size Issues for Conjoint Analysis Studies. Research Paper Series.
Pearmain, D., Gleave, S.D., 1991. Stated Preference Techniques: A Guide to Practice. Steer Davies Gleave.
Rose, J.M., Bliemer, M.C.J., 2013. Sample size requirements for stated choice experiments. Transportation 40, 1021–1041.
Sándor, Z., Franses, P.H., 2009. Consumer price evaluations through choice experiments. J. Appl. Econometrics 24 (3), 517–535.
Swait, J., Adamowicz, W., 2001. Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. Organ. Behav. Hum. Decis. Process. 86 (2), 141–167.
Tang, J., Vandale, W., Weiner, J., Insight, I., 2006. Sample planning for choice-based conjoint models: Our experience. In: Proceedings of the 2006 Conference. Sawtooth Software Inc., Florida, pp. 111–116.
Traets, F., Sanchez, D.G., Vandebroek, M., 2020. Generating optimal designs for discrete choice experiments in R: The idefix package. J. Stat. Softw. 96 (3), 1–41. http://dx.doi.org/10.18637/jss.v096.i03.
Vanniyasingam, T., Daly, C., Jin, X., Zhang, Y., Foster, G., Cunningham, C., Thabane, L., 2018. Investigating the impact of design characteristics on statistical efficiency within discrete choice experiments: A systematic survey. Contemp. Clin. Trials Commun. 10, 17–28. http://dx.doi.org/10.1016/j.conctc.2018.01.002.
Yang, J.C., Johnson, F.R., Kilambi, V., Mohamed, A.F., 2015. Sample size and utility-difference precision in discrete-choice experiments: A meta-simulation approach. J. Choice Model. 16, 50–57. http://dx.doi.org/10.1016/j.jocm.2015.09.001.
Zahran, A., Anderson-Cook, C.M., Myers, R.H., 2003. Fraction of design space to assess prediction capability of response surface designs. J. Qual. Technol. 35 (4), 377–386. http://dx.doi.org/10.1080/00224065.2003.11980235.