

Chapter 3

Experimental Design



Abstract This chapter covers various issues related to the experimental design, a statistical technique at the core of a discrete choice experiment. Specifically, it focuses on the dimensionality of a choice experiment and the statistical techniques used to allocate attribute levels to choice tasks. Among others, the pros and cons of orthogonal designs, optimal orthogonal in the differences designs as well as efficient designs are addressed. The last section shows how a simulation exercise can help to test the appropriateness of the experimental design.

3.1 The Dimensionality of a Choice Experiment

The following five features can characterise the dimensionality of a choice experiment: the number of attributes, the number of levels used to describe the corresponding attribute, the range of the attribute levels, the number of alternatives presented in a choice task and, finally, the number of choice tasks. Considering the dimensions of a DCE is important as trade-offs might exist between their size and what is referred to as response efficiency. Response efficiency, according to Johnson et al. (2013, p. 6), refers to “measurement error resulting from respondents’ inattention to the choice questions or other unobserved, contextual influences”. Therefore, a low response efficiency means that respondents are less likely to identify the alternatives they prefer the most and will reduce choice consistency, i.e. the unexplained part or error term will vary to a greater extent. However, this effect does not take place uniformly for all design dimensions as the literature shows.

Two studies so far have systematically investigated the influence of all five dimensions on respondents’ choices: Caussade et al. (2005) in transportation and Meyerhoff et al. (2015), building on Caussade et al. (2005) and Hensher (2006), in environmental economics. Both studies have used a so-called design-of-designs approach. Other important studies on this topic have been conducted by DeShazo and Fermo (2002), Boxall et al. (2009), Boyle and Özdemir (2009), Rolfe and Bennett (2009), Zhang and Adamowicz (2011), Hess et al. (2012), Czajkowski et al. (2014), and Campbell et al. (2015). Below we look at the various design dimensions separately.

3.1.1 Number of Choice Tasks

People responsible for designing a DCE are often afraid of presenting respondents with too many choice tasks. There are several published papers where it is suggested that presenting respondents “with more than four or six choice tasks” would be too much for them as it would be too complex and respondents would tire when having to respond to numerous tasks. However, the literature does not support this idea. There is, of course, a maximum number of choice tasks an individual is able (and willing) to respond to, but the number of tasks that respondents can answer before becoming fatigued seems to be higher than is often assumed. Hess et al. (2012), investigating different data sets from choice experiments conducted in transportation, argue that concerns about fatigue are probably overstated. Accommodating for scale heterogeneity had little or no impact on substantive models results, and the role of the constants in the models generally decreased. Czajkowski et al. (2014), for example, presented respondents with 26 choice tasks and were not able to identify clear signs of fatigue. Meyerhoff et al. (2015) were also not able to conclude that respondents who faced numerous choice tasks were significantly more likely to drop out of the survey. They presented splits of respondents in their design-of-designs approach with 6, 12, 18 and 24 choice tasks. Also Campbell et al. (2015) could not find strong evidence for fatigue in their study either, respondents were asked to respond to 16 choice tasks. Presenting more choice tasks than originally thought is therefore an option to be considered.

Moreover, a higher number of choice tasks is also crucial when calculating individual-specific WTP values as these conditional values are only meaningful when a sufficient number of choices is available for each respondent (Train 2009, Chap. 11; Sarrias 2020). However, further research would be helpful as the present findings might depend on the specific study contexts or on survey mode. Responding to 16 choice tasks in an online survey might, for example, be different from responding to 16 choice tasks in a paper and pencil survey. In any case, it is important to test prior to the survey whether the intended number of choice tasks can be considered manageable for the average respondent.

3.1.2 Number of Attributes

The studies by Caussade et al. (2005) and Meyerhoff et al. (2015) also suggest that increasing the number of attributes does not affect response efficiency negatively. Caussade et al. (2005) varied the number of attributes from 3 to 6, while Meyerhoff et al. (2015) varied them from 4 to 7. However, both expanded the number of attributes without adding new content. Caussade et al. (2005) presented to a split sample, for example, the attributes “free flow time” and “congestion time” instead of the attribute “total travel time” to increase the number of attributes. Meyerhoff et al. (2015) increased the number of attributes by splitting the attribute “overall biodiversity”

into “biodiversity in forests” and “biodiversity in other parts of the landscape”, for instance. Thus, it is not clear from either study whether this approach of expanding attributes is the reason why negative effects are not found with a higher number of attributes. Outcomes might be different when each attribute introduces a new characteristic of the good in question and therefore would clearly increase the amount of information a respondent would have to process. For the selection of attributes, see also Greiner et al. (2014).

3.1.3 Number of Alternatives

A dimension that might be more critical in terms of negative impacts on response efficiency is probably the number of alternatives. Findings by Zhang and Adamowicz (2011) suggest that with a larger number of alternatives the complexity increases. They compared choice tasks with two and choice tasks with three alternatives. They also point out that the increase in complexity might outweigh the benefits from the fact that people who are presented with more alternatives are more likely to find the alternative that matches their preferences best. Boyle and Özdemir (2009) find that respondents were more likely to choose the status quo (SQ) alternative when there were three alternatives on a choice task compared to tasks with two alternatives. This finding is supported by Oehlmann et al. (2017) who found that the number of alternatives has a significant impact on the frequency of status quo choices, i.e. the alternative with a zero price offer describing the current situation. The more alternatives a choice task comprised, the less often the status quo alternative was chosen.

A processing strategy that might be triggered by the number of alternatives is a switch from comparing the overall utility of an alternative to using the levels of the cost attribute as an indicator of quality alone. Meyerhoff et al. (2017) compared the effects of varying the number of choice tasks by comparing results from split samples where respondents faced different numbers of alternatives. In the splits with four and five alternatives, in addition to the status quo alternative, people seem to be more likely to switch to cost as an indicator of quality. In contrast, Czajkowski et al. (2014) observed no differences to WTP estimates when comparing choice tasks with two and three alternatives.

3.1.4 Other Dimensionality Issues

The number of attribute levels and the value range of the levels can have a positive effect on response efficiency and thus, choice consistency but also in identifying potential non-linear relationships for a given attribute. In line with the findings by Caussade et al. (2005), Meyerhoff et al. (2015) found that a higher number of attribute levels seems to impact on choice consistency positively, as does a narrow range of

the level values. In both cases, it is probably easier for respondents to identify the preferred alternative when comparing the set of alternatives presented on a choice task. Also a higher number of attribute levels also makes a level balanced design more likely (see Sect. 3.2).

Another important point to consider is the randomisation of the order of appearance of the choice tasks if the survey mode allows for this to reduce the impact of anchoring (Jacobsen and Thorsen 2010) and to accommodate for scale heterogeneity (see Sect. 6.2). Also note that respondents might react differently to a long sequence of tasks in an online survey compared to a paper and pencil survey, so knowing the survey mode when deciding on the design dimensions is beneficial.

Regarding attribute non-attendance (Sect. 6.5), Weller et al. (2014) investigated whether stated or inferred attribute non-attendance are linked to the dimensions of the DCE. Overall, their results indicated only a weak relationship between attribute non-attendance and the design dimensions. They suggest, however, that a higher degree of non-attendance might take place when the number of alternatives and choice sets increases; more evidence is needed to draw stronger conclusions here.

A recommendation made by Zhang and Adamowicz (2011) is supported here. If you can afford another split in your survey design, you may consider employing choice tasks with only two alternatives that are said to perform better concerning incentive compatibility (see Sect. 2.4). Splits with choice tasks with two alternatives provide a yardstick for judging the effects of choice task with more alternatives. Also, if the sample is large enough and the order of appearance is randomised, it is possible to estimate simple models such as the conditional logit using only the responses to the first choice task each respondent faced while checking for potential differences.

An issue that requires further research is the relationship between dimensionality and incentive compatibility (see also Sect. 2.4). Generally, binary choices are seen as incentive compatible, i.e., respondents to this format should theoretically reveal true preferences. Whether this also applies to (a) a sequence of tasks with two alternatives, and (b) to sequences of choice tasks with more than two alternatives is still an open question. Vossler et al. (2012) show that under certain conditions, sequences of binary choice questions are incentive compatible but additional work on the association between the dimensionality of a choice experiment and incentive compatibility would be well received.

3.2 Statistical Design of the Choice Tasks

The purpose of an SP study is to learn about individual preferences. The benefit of using an SP survey is that, in contrast to RP, we can control the choices we present to people. In designing these choice tasks, two criteria are of importance. First, the choices presented to respondents need to be relevant. Second, the informational content (from a statistical point of view) of the design needs to be maximised. We need to present respondents with the trade-offs that provide us the best possible

information about the preferences in the sample of interest (i.e. the coefficients of the utility function). Below, it is assumed that the attributes and the relevant levels are given and have been defined in a stage prior to the experimental design.

Originally, orthogonal designs were applied in DCE. Orthogonal designs ensure that the attribute levels are independent of each other, i.e. have zero correlation. In linear economic models, such as the linear regression model, orthogonal designs are also optimal from a statistical point of view. However, when working with discrete choice models, which are highly non-linear, this equivalence no longer holds. It is important to note that the underlying utility functions may be linear-in-parameters, but the choice probabilities are highly non-linear. A benefit of orthogonal designs is that they remove the correlation across key attributes of interest and thereby allow easy identification of their influence on utility. Moreover, orthogonal designs ensure that (i) every pair of attribute levels appears equally often across all pairs of alternatives and (ii) attribute levels are balanced, i.e. each level occurs the same number of times for each alternative.

Orthogonality, however, does not consider the realism of the choice tasks and often the design includes alternatives that are dominated (e.g. both worse in quality and more expensive). Also, random and orthogonal designs are more robust across modelling assumptions but inherently result in a loss of efficiency (Yao et al. 2015). Hence, alternative design generation strategies were being formulated. One of these strategies is Optimal Orthogonal in the Differences (OOD) designs as introduced by Street et al. (2001, 2005). These *D-optimal* designs still maintain orthogonality, but attributes that are common across alternatives are not allowed to take the same level in the design, hence the term *optimal in the differences*. The Ngene manual (ChoiceMetrics 2018) highlights that OOD designs can only be used for unlabelled experiments and may stimulate certain types of behaviour since specific attributes may influence the entire experiment given that the levels are never the same across alternatives. Due to this nature of OOD designs, efficient designs have developed as a popular alternative. By optimising for a specific utility function, we obtain more information about the parameters of interest from the same amount of choices.

More information typically means obtaining more efficient parameter estimates and generally that implies lower standard errors. However, the efficient design literature makes use of alternative efficiency definitions. That is, different definitions of efficiency have an objective that goes beyond reducing the standard error of the parameter estimates. To make this clearer, we need to trace back to the origin of the standard errors. They are generally obtained from the Hessian (i.e. the matrix of second-order derivatives of the log-likelihood function) evaluated at the estimated values of the parameters. The Hessian summarises all the uncertainty associated with the parameters of interest. The negative inverse of this matrix is also known as the asymptotic covariance (AVC) matrix of parameter estimates and the square root of the diagonal terms gives us our standard errors of interest. The off-diagonal elements capture the extent to which alternative parameters can be identified independently from each other. The latter is crucial information since reducing the standard error on one parameter may mean we may no longer be able to separate that specific effect from other attributes in the SP study.

In short, we want to minimise the uncertainty, or maximise the informational content, in our experiment as summarised by the Fisher information matrix. Maximising something, however, requires a unique number and not a matrix. Hence, we need to reduce the dimensionality of the Hessian to a single number and that is where the efficient design *alphabet soup* comes into play (Olsen and Meyerhoff 2017).

The most widely used efficiency measure is the D-error, where alternative designs are compared based on the determinant of the AVC matrix. A D-efficient design is the design that has a sufficiently low D-error. Note that it is often impossible to find the D-optimal design, which has the lowest possible D-error, due to the large number of possible design combinations. By focusing on the determinant, it does not solely focus on minimising the standard errors, but also takes into account the degree of correlation between parameters. The D-error can also be directly related to the measure of information in the Fisher information matrix through the eigenvectors, hence explaining the popularity of this measure. Software packages, such as Ngene (ChoiceMetrics 2018), also allow us to find efficient designs using alternative efficiency measures:

- (a) A-efficiency: this efficiency measure minimises the trace of the AVC matrix and thereby only looks at the variances (standard errors) and not the covariances between parameters estimates. It is important for this measure to work effectively that all parameters are of comparable scale.
- (b) C-efficiency: this efficiency measure works particularly well when interested in WTP measures since it focuses on minimising the variances (standard errors) of parameter ratios.
- (c) D-efficiency minimises the determinant of the Hessian. Thus, it tries to minimise the standard errors on the diagonal, while at the same time controlling for the degree of correlation between parameter estimates. The D-efficiency criterion is the most commonly used criterion in the literature.
- (d) S-efficiency: this efficiency criterion finds its origin in the *t*-value (ratio of the parameter over its standard error). It aims to identify the number of repetitions in the design that are needed for a parameter to be significant. S-efficient designs spread the amount of information across the parameters of interest and hence minimises the number of repetitions needed to obtain significant parameter estimates for all parameters. The S-statistic is merely a lower bound, since the optimisation assumes that respondents act according to the specified prior parameter values.

An detailed description of the alternative design measures and the theory of efficient design is given in the Ngene manual (ChoiceMetrics 2018). It should be noted that all efficiency criteria make use of the AVC matrix, which inherently depends on the parameters of the model. More explicitly, the AVC matrix of the multinomial logit (MNL) model is a function of the parameters of the model. This explains the requirement of efficient designs to define prior parameter values when generating the design. As such, the design will be optimised for these specific parameter values and is therefore optimised locally. If preferences in society differ, it is therefore

not guaranteed that this will be the best design. Alternative strategies can therefore be employed. First, it is always good practice to base prior parameters on existing values in the literature. Second, it is also common practice to generate an initial design based on non-efficient design criteria (random designs, or orthogonal designs). This non-optimal design then serves as the basis in a pre-test from which a set of prior values can then be elicited. However, it needs to be ensured that the sample size of the pre-test is sufficiently large to make useful inferences about the parameters of interest.

Even after employing these strategies, the researcher is typically left with a significant degree of uncertainty about the parameters of interest. To optimise the design over a larger region of parameter estimates one typically reverts to Bayesian designs. The terminology for Bayesian designs is rather unfortunate, since the design criterion is still based on the AVC matrix which plays no role of interest in Bayesian estimation. Nevertheless, the terminology does capture that the parameters of interest are inherently uncertain. The researcher is therefore requested to specify a prior density (e.g. normal or uniform distribution) describing the possible range and likelihood for the potential parameter values (Bliemer and Collins 2016). The design generation then optimises the design by taking a weighted average of the design criterion over all possible parameter values. A direct result of optimising over a wider range of parameter values is that the design is more generic and is thereby likely to lose some efficiency. However, this would only be the case when we accurately know our parameters of interest. Bayesian designs can therefore be labelled as good practice. A general guideline here is that the less known about the parameter estimates of interest, the wider the range should be of parameter values specified for the Bayesian design to reflect this uncertainty.

The AVC matrix does not only depend on the parameters of interest, but also on our assumption about the error term and the functional form of the utility function. Van Cranenburgh et al. (2018), for example, illustrate that designs generated for a RUM decision criterion may not be overly suited to identify choices based on a Random Regret Minimisation (RRM) decision rule. Similarly, Ngene (ChoiceMetrics 2018) allows us to generate designs for non-MNL models, such as nested logit and MXL. Indeed, such models are associated with a much more complicated likelihood function and thus a definition of the Hessian, but the underlying principles of generating efficient designs are not affected. The challenge, however, is that a priori we typically do not know which models we will estimate. Moreover, unlike Bayesian efficient designs, there are currently no design algorithms that allow optimisation of the design over a range of model specifications. As such, it is good practice to generate the design for the most generic model possible (typically the MXL). Generating mixed logit designs takes much longer and is therefore often avoided despite being good practice. An alternative is again to use random or orthogonal designs which are more robust across modelling assumptions but inherently result in a loss of efficiency. In the end, the researcher should be reminded that variations in the attribute levels is of most importance and that efficient designs are only aimed at obtaining more information from the same amount of choices *for a set of given modelling assumptions*.

Recently, the focus in the literature has been on the generation of efficient designs. Statistical efficiency is, however, not the panacea and only criteria that determine the quality of the design. An efficient design is optimised for a given model and there are numerous reasons why that model may be misspecified and hence it would not be appropriate to characterise the response behaviour. Accordingly, it is considered good practice to have a larger number of choice tasks to better cover the space of potential attribute level combinations.

Finally, most experimental designs are only based on main effects and do not consider interaction effects between parameters. As an analyst, when we wish to learn about two-way interaction effects (i.e. how combinations of attributes and their levels influence utility) this requires presenting specific combinations of attribute levels. These requirements can be accommodated in both orthogonal and efficient designs relatively easily. However, to empirically identify interaction effects typically significantly larger sample sizes are required as opposed to identifying main effects. To see this, one can easily compare the S-efficiency statistic across designs (not including interaction effects).

In summary, practitioners should bear in mind that the key to obtaining informative results is presenting respondents with different trade-offs. Hence, the more attribute levels and the more choice tasks the better. Using blocking across respondents to obtain more versions of the design to learn more about preferences across respondents may also be recommended. Alternatively, tasks can be randomly assigned to respondents, especially when the overall number of choice tasks is rather large. Also, when developing surveys start off with simple orthogonal designs or random designs and use the result from the pilot for updating the priors. Finally, convention so far states that MNL-based efficient designs perform well and not much worse compared to the designs optimised for more advanced models (Bliemer and Rose 2010, 2011).

3.3 Checking Your Statistical Design

The so-called right-hand side matrix in a linear regression is formed by the explanatory variables. In a discrete choice model, this matrix is defined by the variables included in V_{njt} in Eq. (1.3) that can be alternative specific constants, attributes, individual-specific variables or their interactions. The right-hand side matrix of discrete choice models plays a crucial role in parameter identification and the precision of their estimation. As described above, the right-hand side matrix in SP data sets is usually set by the experimental design. A high number of attributes, and/or attribute levels, can make the search for a convenient experimental design a tricky task. The literature on experimental designs (Street and Burgess 2007; Louviere and Lancsar 2009; ChoiceMetrics 2018) describes how to generate them, how to analyse their properties and efficiency or how to block them. Nevertheless, in the applied literature, not sufficient attention is usually paid to all these steps and they are usually not sufficiently described. Moreover, sometimes the coding used in the experimental design has been changed in the econometric analysis. For example, efficient designs

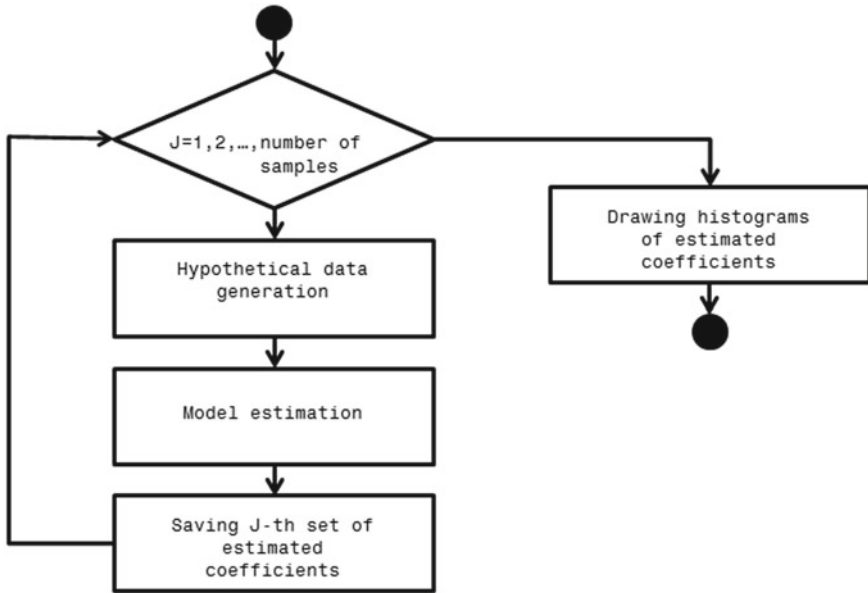


Fig. 3.1 Flowchart of a simulation exercise

with attribute levels specified as continuous (e.g. 1, 2, 3, 4) are coded as categorical after the data were collected. This categorical coding can be inappropriate for parameter identification.

The appropriateness of an experimental design or, generally speaking, the appropriateness of the right-hand side matrix of discrete choice models can be easily checked by a simulation exercise presented in Fig. 3.1.

This check is based on the generation of numerous hypothetical data sets based on the generated (SP data) or collected (revealed preference (RP) data) right-hand side matrix. The hypothetical data sets are generated by setting the values of the parameters to a specific value assuming that these are the true population values and generating specific values of the error components. In each iteration, a hypothetical data set is used for a model estimation and the set of estimated parameters is saved.

Post-analysis of the empirical distribution of all parameters can reveal whether the right-hand side matrix allows for an unbiased estimation of all the parameters, as the true population parameters are known. This simple simulation exercise should always be carried out both in RP and in SP studies. In RP studies, it allows us to check whether the variation of the collected attribute levels is sufficient to identify all the parameters correctly. In SP studies, it allows us to check the appropriateness of the generated experimental design as well as the expected distribution of the parameter estimates.

For example, imagine we want to analyse the appropriateness of the following experimental design

alt1.attr1	alt1.attr2	alt2.attr1	alt2.attr2	alt3.attr1	alt3.attr2
1	3	3	5	9	9
7	1	7	7	5	5
7	9	5	1	5	9
1	3	9	1	7	7
5	9	3	9	7	1
9	5	1	7	1	3
3	7	9	3	1	5
5	1	7	9	3	3
9	7	1	3	3	7
3	5	5	5	9	1

corresponding to a one choice-occasion with three alternatives and two attributes defined according to the Eq. (1.4), as

$$\begin{aligned}
 U_{n1} &= ASC_1 + \beta_1 \text{attr}_{n1} + \beta_2 \text{attr}_{n2} + \varepsilon_{n1} \\
 U_{n2} &= ASC_2 + \beta_1 \text{attr}_{n2} + \beta_2 \text{attr}_{n2} + \varepsilon_{n2} \\
 U_{n3} &= \beta_1 \text{attr}_{n3} + \beta_2 \text{attr}_{n3} + \varepsilon_{n3}
 \end{aligned}$$

Subsequently, we assume that the following values of the parameters are population values

$$\begin{aligned}
 U_{n1} &= 0.5 + 0.1 \text{attr}_{n1} - 0.1 \text{attr}_{n2} + \varepsilon_{n1} \\
 U_{n2} &= 0.5 + 0.1 \text{attr}_{n2} - 0.1 \text{attr}_{n2} + \varepsilon_{n2} \\
 U_{n3} &= 0.1 \text{attr}_{n3} - 0.1 \text{attr}_{n3} + \varepsilon_{n3}
 \end{aligned}$$

and generate, for example, 5,000 times three sets of Gumbel-distributed errors ε_{n1} , ε_{n2} and ε_{n3} for a specific sample size. Using these sets of errors, the above-presented design and the assumed coefficient values, we can generate 5,000 utilities U_{n1} , U_{n2} and U_{n3} , and therefore, 5,000 hypothetical choices. Then, we can estimate 5,000 times a MNL model and draw histograms of these estimates for each parameter. This is how we can analyse, for example, the impact of the number of observations on the precision of the estimates based on the generated design.

Figure 3.2 presents histograms of 5,000 estimations of the four above-defined coefficients. The first column in Fig. 3.2 shows the histograms for 100 observations and the second row for 400 observations. This example shows, in a very simple and graphic way, two well-known findings. Firstly, the estimation of the coefficients in our MNL model by maximum likelihood is consistent, because the spread of estimations in the second column in Fig. 3.2 is narrower. Secondly, focusing on the x-axis of the histograms, the precision of the estimations of the alternative specific constants is in our case worse than the precision of the attribute coefficients. Please note that

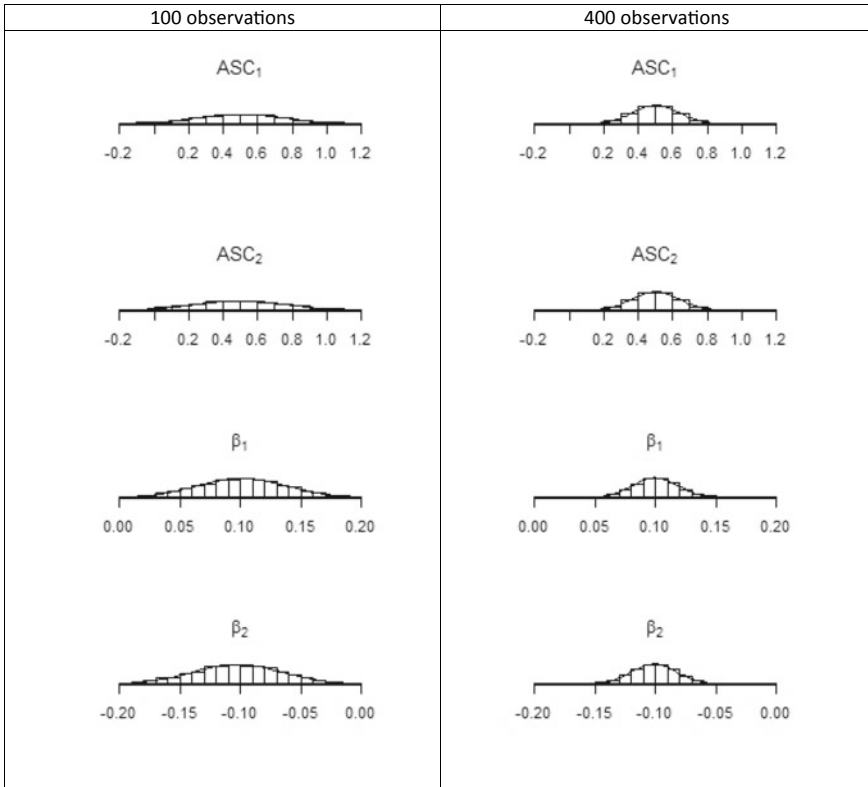


Fig. 3.2 Histograms

all histograms are centred on the assumed population value ($ASC_1 = 0.5$, $ASC_2 = 0.5$, $\beta_1 = 0.1$, $\beta_2 = -0.1$) confirming the appropriateness of the experimental design in providing unbiased estimates of the population parameter values.

References

Bliemer MCJ, Collins AT (2016) On determining priors for the generation of efficient stated choice experimental designs. *J Choice Model* 21:10–14. <https://doi.org/10.1016/j.jocm.2016.03.001>

Bliemer MCJ, Rose JM (2010) Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transp Res Part B Method* 44:720–734. <https://doi.org/10.1016/j.trb.2009.12.004>

Bliemer MCJ, Rose JM (2011) Experimental design influences on stated choice outputs: an empirical study in air travel choice. *Transp Res Part A Policy and Practice* 45:63–79. <https://doi.org/10.1016/j.tra.2010.09.003>

- Boxall P, Adamowicz WL, Moon A (2009) Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement. *Aust J Agric Resource Econ* 53:503–519. <https://doi.org/10.1111/j.1467-8489.2009.00469.x>
- Boyle KJ, Özdemir S (2009) Convergent validity of attribute-based, choice questions in stated-preference studies. *Environ Resource Econ* 42:247–264. <https://doi.org/10.1007/s10640-008-9233-9>
- Campbell D, Boeri M, Doherty E, George Hutchinson W (2015) Learning, fatigue and preference formation in discrete choice experiments. *J Econ Behav Organ* 119:345–363. <https://doi.org/10.1016/j.jebo.2015.08.018>
- Caussade S, Ortúzar J de D, Rizzi LI, Hensher DA (2005) Assessing the influence of design dimensions on stated choice experiment estimates. *Transp Res Part B Methodol* 39:621–640. <https://doi.org/10.1016/j.trb.2004.07.006>
- ChoiceMetrics (2018) Ngene 1.2 user manual & reference guide. Australia
- Czajkowski M, Giergiczyński M, Greene WH (2014) Learning and fatigue effects revisited: investigating the effects of accounting for unobservable preference and scale heterogeneity. *Land Econ* 90:324–351. <https://doi.org/10.3368/le.90.2.324>
- Deshazo JR, Fermo G (2002) Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *J Environ Econ Manage* 44:123–143
- Greiner R, Bliemer M, Ballweg J (2014) Design considerations of a choice experiment to estimate likely participation by north Australian pastoralists in contractual biodiversity conservation. *J Choice Model* 10:34–45. <https://doi.org/10.1016/j.joem.2014.01.002>
- Hensher DA (2006) Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment. *Environ Resource Econ* 34:7–44. <https://doi.org/10.1007/s10640-005-3782-y>
- Hess S, Stathopoulos A, Daly A (2012) Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. *Transportation* 39:565–591. <https://doi.org/10.1007/s11116-011-9365-6>
- Jacobsen JB, Thorsen BJ (2010) Preferences for site and environmental functions when selecting forthcoming national parks. *Ecol Econ* 69:1532–1544. <https://doi.org/10.1016/j.ecolecon.2010.02.013>
- Louvière JJ, Lancsar E (2009) Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Econ Policy Law* 4:527–546. <https://doi.org/10.1017/S1744133109990193>
- Meyerhoff J, Mariel P, Bertram C, Rehdanz K (2017) Matching preferences or changing them? The influence of the number of choice alternatives. In: 23rd Annual Conference of the European Association of Environmental and Resource Economists. Athens, Greece
- Meyerhoff J, Oehlmann M, Weller P (2015) The Influence of design dimensions on stated choices in an environmental context. *Environ Resource Econ* 61:385–407. <https://doi.org/10.1007/s10640-014-9797-5>
- Oehlmann M, Meyerhoff J, Mariel P, Weller P (2017) Uncovering context-induced status quo effects in choice experiments. *J Environ Econ Manage* 81:59–73. <https://doi.org/10.1016/j.jeem.2016.09.002>
- Olsen SB, Meyerhoff J (2017) Will the alphabet soup of design criteria affect discrete choice experiment results? *Eur Rev Agric Econ* 44:309–336. <https://doi.org/10.1093/erae/jbw014>
- Reed Johnson F, Lancsar E, Marshall D et al (2013) Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint Analysis experimental design good research Practices task force. *Value in Health* 16:3–13. <https://doi.org/10.1016/j.jval.2012.08.2223>
- Rolfe J, Bennett J (2009) The impact of offering two versus three alternatives in choice modelling experiments. *Ecol Econ* 68:1140–1148. <https://doi.org/10.1016/j.ecolecon.2008.08.007>
- Sarrias M (2020) Individual-specific posterior distributions from Mixed Logit models: properties, limitations and diagnostic checks. *J Choice Model* 100224. <https://doi.org/10.1016/j.joem.2020.100224>

- Street DJ, Bunch DS, Moore BJ (2001) Optimal designs for 2k paired comparison experiments. *Commun Stat Theory Methods* 30:2149–2171. <https://doi.org/10.1081/STA-100106068>
- Street DJ, Burgess L (2007) *The construction of optimal stated choice experiments: theory and methods*. Wiley, United States
- Street DJ, Burgess L, Louviere JJ (2005) Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments. *Int J Res Mark* 22:459–470. <https://doi.org/10.1016/j.ijresmar.2005.09.003>
- Train K (2009) *Discrete choice methods with simulation*, 2nd edn. Cambridge University Press, New York
- van Cranenburgh S, Rose JM, Chorus CG (2018) On the robustness of efficient experimental designs towards the underlying decision rule. *Transp Res Part A Policy Pract* 109:50–64. <https://doi.org/10.1016/j.tra.2018.01.001>
- Vossler CA, Doyon M, Rondeau D (2012) Truth in consequentiality: theory and field evidence on discrete choice experiments. *Am Econ J Microecon* 4:145–171. <https://doi.org/10.1257/mic.4.4.145>
- Weller P, Oehlmann M, Mariel P, Meyerhoff J (2014) Stated and inferred attribute non-attendance in a design of designs approach. *J Choice Model* 11:43–56. <https://doi.org/10.1016/j.jocm.2014.04.002>
- Yao RT, Scarpa R, Rose JM, Turner JA (2015) Experimental design criteria and their behavioural efficiency: an evaluation in the field. *Environ Resource Econ* 62:433–455. <https://doi.org/10.1007/s10640-014-9823-7>
- Zhang J, Adamowicz WL (2011) Unraveling the choice format effect: a context-dependent random utility model. *Land Economics* 87:730–743. <https://doi.org/10.3368/le.87.4.730>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

