Choice experiments in non-market value analysis: some methodological issues

Dieter Koemle

The Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany, and

Xiaohua Yu

Department of Agricultural Economics and Rural Development, University of Göttingen, Göttingen, Germany

Abstract

Purpose – This paper reviews the current literature on theoretical and methodological issues in discrete choice experiments, which have been widely used in non-market value analysis, such as elicitation of residents' attitudes toward recreation or biodiversity conservation of forests.

Design/methodology/approach – We review the literature, and attribute the possible biases in choice experiments to theoretical and empirical aspects. Particularly, we introduce regret minimization as an alternative to random utility theory and sheds light on incentive compatibility, status quo, attributes non-attendance, cognitive load, experimental design, survey methods, estimation strategies and other issues.

Findings – The practitioners should pay attention to many issues when carrying out choice experiments in order to avoid possible biases. Many alternatives in theoretical foundations, experimental designs, estimation strategies and even explanations should be taken into account in practice in order to obtain robust results.

Originality/value – The paper summarizes the recent developments in methodological and empirical issues of choice experiments and points out the pitfalls and future directions both theoretically and empirically.

Keywords Choice experiment, Methodological issues, Order effects, Experimental design, Incentive compatibility

Paper type Conceptual paper

1. Introduction

Choice experiments have been used for a long time to estimate consumer preferences and predict consumer behavior in market (Gao and Schroeder, 2009; Louviere and Hensher, 1982; Lusk and Schroeder, 2004) and non-market valuation studies (Adamowicz *et al.*, 1998; Boxall *et al.*, 1996; Morey *et al.*, 2002). Forests have ecological multifunction and the non-market values cannot be easily elicited. Ample applications of choice experiments have been carried out for studying residents' attitudes toward recreation (Sælen and Ericson, 2013; Juutinenab *et al.*, 2014), carbon sequestration, biodiversity conservation and ecological services (Baranzini *et al.*, 2012) and other conservation values of forests (Cerda, 2006; Cerda *et al.*, 2014).

A choice experiment is a survey approach designed to elicit consumer preferences based on hypothetical markets. Respondents are required to choose between multiple public or private goods. This choice is expected by the researcher to occur by trading of the individual

attributes of the different goods available, and choosing the good (or alternative) that provides the most utility. This approach to consumer behavior was first developed by Lancaster (1966), who states that the utility from a good is not derived from the good itself, but from its individual attributes. From a series of observed choices, a researcher then tries to infer the latent utility function. Traditionally, McFadden's (1974) random utility approach is used to describe the utility gained from a certain alternative on the basis of the attributes, utility weights for each attribute and a random error term to make the estimation of the utility weights feasible. Finally, the estimated model can be used for welfare estimation or market share predictions.

However, researchers are faced with a number of choices when designing a choice experiment. The initial step in designing a choice experiment is the development of an attribute list. The number and type of attributes (either quantitative or qualitative) critically depend on the decision-making context, and attributes need to be thoroughly tested. For an economic valuation study, it is essential that one attribute capturing the cost of the alternative is included. Next, levels must be assigned to each attribute, where great care must be given to realism and local nuances. Depending on the size of each alternative (in terms of number of attributes), the researcher must decide on how many alternatives to include in a single choice set, and whether an "opt-out"-option should be included. In market good valuation studies, these alternatives usually describe the "would not buy any" option in a choice set. When choosing among non-market goods such as environmental amenities, this option is often considered as a "status quo" option, which simply describes the state that the respondent is currently in.

After the researcher decided on the number of, attributes, levels and alternatives, the next step becomes developing an experimental design. Full factorial or orthogonal fractional factorials, D-optimal designs or Bayesian designs have been proposed in the literature. When the design has been created and prepared such that the cognitive burden to the respondent is low enough to create reliable responses, supporting questions of the questionnaire can be developed. These include socio-demographic questions, but can also aim at attitudes, behavioral aspects or attribute attendance. Debriefing questions may help investigate possible reliability issues of the model estimates in later stages of the analysis.

To elicit preferences, one or several survey modes must be chosen. Examples include mail or online surveys, or mixed-mode approaches. Respondents can be contacted via professional survey companies, or researchers may prefer to draw random samples themselves. Hidden populations might even be contacted via Internet forums or via snowball sampling.

After the sample has been collected, model estimation is the next step. Here, the researcher has to decide which type of model should be estimated. The standard model is the multinomial logit (MNL) model. However, in recent years, a number of other models have been developed which avoid some of the restrictive assumptions like the independence of irrelevant alternatives (IIA) assumption or the preference homogeneity assumption of MNL model. The random parameters logit (RPL) model assumes some form of distribution of the parameters and therefore allows for preference heterogeneity across the sample. Latent class models allow to model choices of discretely distributed, "latent" respondent types and computationally separate respondents into different classes. While the traditional underlying model in the analysis of choices has been the random utility model, recent developments have incorporated regret theory into choice models (Chorus et al., 2008). Both models can be estimated using the methods described above, but differences in model interpretation and the underlying decision framework make a closer look interesting and necessary.

Finally, results can be used to estimate the benefits of policies across the target population, or the willingness to pay for new products. Further, preference weights can be used to predict consumer behavior in specific scenarios (see Hensher et al. (2005); Louviere et al. (2000)).

While choice experiments have already been used for decades (see e.g. Hanley *et al.* (1998); Hanley *et al.* (2001) and Hoyos (2010) for comprehensive reviews of the application to environmental policy choices), the method has rapidly developed in theoretical and methodological issues, attempting to make the method better fit the framework of economic theory and human decision processes. Therefore, there is a call to summarize these advances. In particular, we focus on, but do not restrict ourselves to studies regarding the valuation of non-market goods. We compare the findings in the different steps of conducting a choice experiment and conclude with some general recommendations. An overview over the publications discussed in this paper and their main innovations is given in Table 1.

2. Method
In this paper, we systematically review the literature on important issues in choice experiments. We used the scientific search engines Google Scholar, ScienceDirect and PubMed. Our primary search terms included choice experiment, choice issues, attribute processing, regret theory, random regret model, status quo option and incentive compatibility. Secondary search terms included order effects, experimental design, efficient design, pivot design, endogeneity, welfare effects, willingness to pay, qualitative methods and attribute design.

3. Theoretical foundations
In this section, we present some theoretical issues regarding the design and analysis of choice experiments. First, we introduce alternative choice rules by discussing the random regret model developed by Chorus *et al.* (2008). Then, we move on to the issue of incentive compatibility.

3.1 Departures from utility theory
The standard model to analyze discrete choice experiments has been McFadden's (1974) random utility model (RUM). In a choice setting, a respondent i is expected to maximize his utility u_i, which is composed of a deterministic, observable part v_i and a stochastic unobservable part ε_i. Assumptions about the distribution of this error term allow estimating the deterministic utility function with some form of binary econometric framework (see Train (2009) for details). However, in recent years, alternative decision models have been suggested to analyze choices, in particular regret theory (Bell, 1982; Fishburn, 1982; Loomes and Sugden, 1982). According to Zeelenberg (1999, p. 326), regret is "the negative, cognitively based emotion that we experience when realizing or imagining that our present situation would have been better had we acted differently". Applications of regret theory in choice modeling include for example Chorus *et al.* (2008); Boeri *et al.* (2014); Hess and Stathopoulos (2013) or Thiene *et al.* (2012). A complete overview of the regret model and its econometric application as Random Regret Model (RRM) is provided by Chorus (2012). In short, instead of maximizing (expected) utility, respondents are expected to minimize their (anticipated) regret from the non-chosen alternatives. As Chorus *et al.* (2008, p. 15) point out, the two decision paradigms lead to very different outcomes; while utility maximizers prefer alternatives that perform well on most attributes, regret minimizers choose alternatives which perform reasonably well on all attributes. An intuitively appealing form of the regret function (Chorus, 2012, p. 8) is

$$R = \max\{0, [\beta_m(x_{jm} - x_{im})]\} \tag{1}$$

where *(x_jm−x_im)* describes the difference in levels of attribute m between alternatives i and j, and β_m is interpreted as attribute m's potential contribution to the regret function. As is clear

Author (Year)	Journal	Innovation
Regret theory		
Chorus *et al.* (2008)	Transportation Research Part B: Methodological	Regret theory in discrete choice
Thiene *et al.* (2012)	Environmental and Resource Economics	RRM vs. RUM comparison in environmental economics
Hess and Stathopoulos (2013)	Journal of Choice Modeling	Mixing RRM and RUM models
Boeri *et al.* (2013)	Conference-paper at the International Choice Modeling Conference	Monte Carlo comparison of RRM and RUM
Boeri *et al.* (2014)	Transportation Research Part A: Policy and Practice	Probabilistic segmentation of respondents into utility maximizers and regret minimizers
Chorus *et al.* (2014)	Journal of Business Research	Literature review comparing RRM and RUM estimates
Incentive compatibility		
Carson and Groves (2007)	Environmental and Resource Economics	Theoretical treatment of incentive compatibility in stated preference discrete choice
Vossler *et al.* (2012)	American Economic Journal: Microeconomics	Consequentiality as a main factor in incentive compatibility
Hensher (2010a)	Transportation Research Part B: Methodological	Review of sources of hypothetical bias in stated preference studies
Opt-out and don't know		
Boxall *et al.* (2009)	Australian Journal of Agricultural and Resource Economics	Increasing complexity results in increasing opt-outs
Meyerhoff and Liebe (2009)	Land Economics	Attitudinal and socio-demographic influences on opt-out behavior
Lanz and Provins (2012)	CEPE Working Paper Series, ETH Zurich	Socio-demographic influences on opt-out behavior and serial nonparticipation
Von Haefen *et al.* (2005)	American Journal of Agricultural Economics	Hurdle model for serial nonparticipation
Balcombe and Fraser (2011)	European Review of Agricultural Economics	General model for "do not know" responses in choice experiments
Attribute processing and ANA		
Hensher (2007)	Chapter in Kanninen (2007)	Theoretical exposition on different attribute processing strategies, influence of complexity on attribute processing
Hensher (2010b)	Chapter in Proceedings of the International Choice Modeling Conference 2010	Dempster–Shafer belief functions to assess processing strategy, attribute non-attendance, attribute aggregation
Mariel *et al.* (2012)	Conference Paper at the European Association of Environmental Resource Economists	Compare stated and inferred methods to detect attribute non-attendance
Alemu *et al.* (2013)	Environmental and Resource Economics	Investigate reasons for attribute non-attendance
Colombo and Glenk (2013)	Journal of Environmental Planning and Management	Consider attribute non-attendance and alternative non-attendance due to unacceptable attributes
Scarpa *et al.* (2009)	European Review of Agricultural Economics	Develop a latent class and a Bayesian approach to account for attribute non-attendance

Table 1.
Summary of studies examined in this review paper

(continued)

Author (Year)	Journal	Innovation
Hensher *et al.* (2012)	Transportation	Develop a latent class approach to attribute non-attendance with constrained parameters across classes
Puckett and Hensher (2008)	Transportation Research Part E: Logistics and Transportation Review	Adapt estimation for rationally adaptive behavior including adding-up and ignoring attributes using follow-up questions
Kravchenko (2014)	Journal of Choice Modeling	Monte Carlo investigation of effects of attribute non-attendance on parameter estimates
Quan *et al.* (2018)	Agribusiness	Compared attribute non-attendance and full set of choices for WTP for food safety
Order effects		
Day *et al.* (2012)	Journal of Environmental Economics and Management	Empirically testing for various types of order effects
Scheufele and Bennett (2012)	Environmental and Resource Economics	Strategic responses and changes in cost sensitivity along a series of choice sets
McNair *et al.* (2011)	Resource and Energy Economics	Difference in WTP between single and multiple choice sets
Meyerhoff and Glenk (2013)	Working Papers on Management in Environmental Planning, TU Berlin	Instruction choice sets may induce starting point bias
Choice set design and attribute selection		
Coast *et al.* (2012)	Health Economics	Use of qualitative methods in attribute selection, recommendations for reporting the design process
Abiiro *et al.* (2014)	BMC Health Services Research	Detailed description of attribute selection process
Michaels-Igbokwe *et al.* (2014)	Social Science and Medicine	Use of decision mapping processes for attribute selection
Kløjgaard *et al.* (2012)	Journal of Choice Modeling	Description of qualitative process for attribute selection, including observational fieldwork and key informant interviews
Experimental design		
Sándor and Wedel (2001)	Journal of Marketing Research	Bayesian design procedure incorporating managers' beliefs about future market shares into priors
Bliemer *et al.* (2009)	Transportation Research Part B: Methodological	Efficient experimental design for nested logit models
Bliemer and Rose (2010)	Transportation Research Part B: Methodological	Efficient experimental design for random parameters logit models
Ferrini and Scarpa (2007)	Journal of Environmental Economics and Management	Monte Carlo investigation of parameter estimates using designs with vs. without prior information
Gao *et al.* (2010)	Agricultural Economics	Monte Carlo investigations of parameter estimates using different design types and various numbers of attributes and levels
Rose *et al.* (2008)	Transportation Research Part B: Methodological	Pivot designs in computer-aided discrete choice experiments
Survey mode, sampling		
Olsen (2009)	Environmental and Resource Economics	Comparison between mail and Internet survey in choice experiment

(*continued*)

Table 1.

Author (Year)	Journal	Innovation
Estimation strategy, endogeneity		
Walker *et al.* (2011)	Transportation Research Part A: Policy and Practice	BLP approach to treat endogeneity in transportation choice model
Petrin and Train (2010)	Journal of Marketing Research	Control function approach to revealed preference data
Guevara and Ben-Akiva (2010)	Chapter in proceedings of the International Choice Modeling Conference 2010	Use control function method and show link between control functions and latent variables
Guevara and Polanco (2013)	Paper presented at the International Choice Modeling Conference 2013	Use of a multiple indicator solution to correct for endogeneity
Guevara and Ben-Akiva (2012)	Transportation Science	Scale factor correction for models estimated by the control function method

Table 1.

from this definition, the value of the regret function cannot be negative, meaning that if the attribute of the chosen alternative is already better than the non-chosen alternative, regret from this attribute is zero. However, this particular functional form has a discontinuity at 0, which makes it difficult to estimate. Chorus (2012) therefore proposes to approximate the regret function in the following way and to add an IID random error term ε to form the random regret model:

$$RR_i = R_i + \varepsilon_i = \sum_{j \neq i} \sum_m \ln\left(1 + \exp[\beta_m(x_{jm} - x_{im})]\right) + \varepsilon_i \qquad (2)$$

One advantage of regret minimization is the fact that compared to the linear specification of the random utility model, the attributes of the regret model are only semi-compensatory, i.e. do not serve as perfect substitutes. In addition, the model has been fully generalized to estimation of choices under uncertainty (Chorus *et al.*, 2008), however, difficulties arise in the estimation of welfare effects. While the random utility model is deeply rooted in microeconomic welfare theory, welfare measures based on regret theory are just currently being developed (Boeri *et al.*, 2014). In addition, Boeri *et al.* (2014) and Hess and Stathopoulos (2013) show approaches to estimate the proportion of respondents which are utility maximizers and those who minimize regret. Monte Carlo simulations by Boeri *et al.* (2013) indicated that the wrong decision model can lead to significant bias in the estimated parameters.

Chorus *et al.* (2014) review the literature and compare RRM and RUM in 21 studies with regard to (1) model fit, (2) predictive performance and (3) managerial implications. By applying the Ben-Akiva and Swait (1986) test for non-nested models, they look for statistically significant differences in model fit and find that contextual differences matter with regard to which model fits the data better. In general, they find that for important or difficult decisions, such as which car to buy or which policy to choose, the RRM framework fits best, while decisions in leisure activities or travel choice were best modeled by the RUM. With regard to predictive performance and external validity, the RRM was found to perform significantly better than the RUM, however, differences were generally small. Finally, different models were also found to influence managerial implications, for example differences in elasticities and predicted market shares. Chorus *et al.* (2014) conclude that the choice between RUM and RRM should be made on the basis of where each model performs better in terms of model fit and predictive power. Alternatively, researchers may opt for a hybrid model, combining utility maximization and random regret minimization either arbitrarily or within a latent class framework (Boeri *et al.*, 2014; Hess and Stathopoulos, 2013).

However, no matter which decision rule is applied, valid results critically depend on whether respondents reveal their true preference in a questionnaire, and whether the responses are influenced by the structure and mode of the questions being asked. Recent findings on incentive compatibility and order effects are therefore described in the next sections.

3.2 Incentive compatibility

"An allocation mechanism or institution is said to be incentive compatible when its rules provide individuals with incentives to truthfully and fully reveal their preferences" (Harrison, 2007, p. 67).

The difference between hypothetical and actual WTP, known as hypothetical bias, has been the subject of several studies (e.g. Hensher, 2010; Murphy *et al.*, 2005; Yu *et al.*, 2016). While there have been some attempts to explain causes for hypothetical bias, it still lacks a general theory (Murphy *et al.*, 2005). In addition, hypothetical bias can go both ways, depending on the context. For example, Brownstone and Small (2005) and Hensher (2010) found that in transportation research, hypothetical WTP is often lower than actual WTP. On the other hand, when valuing different private or public goods, WTP of the hypothetical scenario has been found to exceed the real WTP when respondents were forced to pay the stated amount for a project (Krawczyk, 2012; Murphy *et al.*, 2005).

In their rigorous theoretical discussion of the incentive compatibility of different choice formats, Carson and Groves (2007) compare single binary choice questions with series of binary and multinomial choice questions. The authors conclude, that in order to be incentive compatible, close attention has to be paid to the good being valued, the choice context and the payment vehicle. For example, valuing a private good without coercive payment might induce a respondent to overstate his WTP in the hypothetical question, if that respondent has at least some probability of gaining positive utility from this good. Overstating own willingness to pay in a non-consequential setting might, in the mind of the respondent, therefore increase the probability of the good being developed. In a non-market good setting, a voluntary payment mechanism might yield similar results. However, if the agency providing the public good can collect the payment coercively, the respondent's incentive to overstate his WTP may be reduced. The statement of true WTP further critically depends on if the respondents perceives the proposed scenario as plausible (meaning the public good could technically be provided at the proposed cost), and how the agency will decide on which good in the choice set will be provided (either by majority rule or some other mechanism).

Vossler *et al.* (2012) conducted an experiment for the valuation of planted trees along roads and rivers. They used four treatments, in which the first three required a real payment, while the fourth treatment leaves the consequentiality of the treatment open. Further, they examined how different policy implementation methods influence choice behavior. Vossler *et al.* (2012) conclude that the notion of consequentiality is far more important then the "real vs. hypothetical" discussion in stated preference applications. Further, their results suggest a 30% increase in WTP for the treatment where no actual payment is defined.

3.3 WTP vs WTA

Practitioners have two choices to elicit non-market values: willingness to pay (WTP) and willing to accept (WTA) (Freeman, 2003). Both theories and empirical evidences show a gap between WTP and WTP (Cerda, 2006; Cerda *et al.*, 2014). The gap can be explained by many factors, such as design methods, respondents' inner attitudes, endowment effects and even legal difference (Freeman, 2003). However, the basic assumptions for WTA and WTP are different and have different legal context. Freemann (2003) points out that the implicit assumption of WTP is that respondents have to accept all policy changes and have to pay for

maintaining the current situation; while WTA assume that respondents have the rights to maintain the current status and are compensated by the policy changes. This assumption has profound implications for experimental design, welfare change and theoretical explanation to the results.

4. Decision process and choice

4.1 Status quo option and "do not know" responses

Most choice experiments include some type of opt-out or status quo option. In the market good context, this could include a "do not buy" option, while for non-market goods, a "I prefer the current situation" is applicable in many cases. While this option adds realism of choice situation, different surveys report "status quo bias" (Samuelson and Zeckhauser, 1988; Zhou *et al.*, 2017) as a possible problem for welfare measurement. This may have various reasons. In the experimental economics literature, status quo bias has been attributed to the endowment effect, preferences for a legitimate alternative, preferences for inaction or to avoid the complexity of a choice task (Boxall *et al.*, 2009). Boxall *et al.* (2009) found evidence that, as the number of choice tasks increases, respondents are more likely to opt out. In addition, their findings indicate that older respondents choose a status quo option consistently more often than younger respondents. Including variables associated with status quo bias significantly changed the levels and the variance of welfare measures associated with some environmental change.

A way to measure a preference for the status quo is to include an alternative specific constant for the status quo. Meyerhoff and Liebe (2009) argue that a significantly positive constant for the status quo could be interpreted either as the average effect of all attributes that were not included or as the utility associated with the status quo option, as suggested by Adamowicz *et al.* (1998). As Meyerhoff and Liebe (2009) demonstrate, the status quo constant can further be interacted with socio-demographic and behavioral characteristics of the respondent. In their choice experiment on sustainable forest management in Lower Saxony, Germany, they found that older, better educated frequent forest users were less, while protest respondents (identified by a number of attitudinal debriefing questions) were more likely to choose status quo. Also, they found some evidence that respondents who perceive the choice task as too complex are more likely to choose status quo. A similar strategy was applied by Lanz and Provins (2012), who also find significant influences of socio-demographic characteristics on status quo choice in the context of water provision in Switzerland. Both studies incorporate attitudinal questions to separate protest responses based on the credibility of the scenario, aversion toward the payment vehicle and the feeling of being provided with insufficient information. Overall, Lanz and Provins (2012) found that all three indicators of protest behavior were significantly increasing the probability of opting out, while variables indicating the perception of the survey (interesting, complicated, educational) were not found to be significant. Interestingly, a more in-depth description of the status quo alternative lead to a significant reduction in status quo responses, all other things equal.

However, in many choice experiments, researchers have to deal with the issue of serial nonparticipation (Von Haefen *et al.*, 2005). In their words, "one form of serial participation is when a respondent always chooses the status quo option" (Von Haefen *et al.*, 2005, p. 1,061). One may argue that the behavioral process guiding serial nonparticipation is different from utility maximization based on attributes, and therefore remove all the respondents engaged in serial nonparticipation from the sample. Von Haefen *et al.* (2005) propose a hurdle model similar to dealing with excess zeros in contingent valuation studies. Lanz and Provins (2012) studied serial nonparticipation based on socio-demographics and protest-attitudes. However, they found no statistically significant evidence that protest attitudes influenced serial nonparticipation, while evidence that satisfaction with the status quo would lead to a higher

probability of serial nonparticipation was statistically significant. The feeling of having been provided with insufficient information however led to a significant higher probability of serial nonparticipation.

While not as frequently used as status quo options, do not know responses can also be introduced into choice sets. Balcombe and Fraser (2011) develop a general framework for the treatment of "do not know" (DK) responses consistent with the nested logit model. Basically, they add the probability of someone giving a DK response, given that actually some other alternative is preferred, to the likelihood function. The likelihood function then becomes

$$f(Y|\beta, \Theta) = \prod_{i=1}^{n} \left(\sum_{k=1}^{J} \theta_{\bullet|k} p_{ik} \right)^{1-\varepsilon_i} \prod_{j=1}^{J} (1 - \theta_{\bullet|j})^{y_{ij}} \prod_{j=1}^{J} p_{ij}^{y_{ij}} \qquad (3)$$

where $\theta_{\bullet|j}$ describes the probability of reporting DK given a preference for alternative j, p_{ij} is the standard logit or probit probability and $\varepsilon_i = 1$ if a preference was reported and zero otherwise. The expression in the first product describes the marginal probability of choosing DK. Balcombe and Fraser (2011) further generalize the model by introducing measures for the similarity between alternatives. They provide three model specifications, one allowing for a constant probability of choosing DK, one where it depends on the similarities between all the alternatives and one where the probability of DK depends on the similarity between the two alternatives that provide the highest utility. Each of these models can be estimated using a specific likelihood function.

4.2 Attribute processing
While the standard assumption in choice experiments is that respondents attend to all attributes equally, evidence has shown that often respondents use simplifying strategies when making their choices. Hensher (2007) cites Payne *et al.* (1992) when summarizing the most important strategies into two broad categories: Attribute based strategies include elimination by aspects, lexicographic choice and majority of confirming dimensions. Alternative-based approaches include weighted additive, satisficing and equal-weight strategies. These strategies differ in the total amount of processing required, and the degree to which processing is consistent or selective across alternatives or attributes (Payne *et al.*, 1992, p. 115).

Hensher (2007) conceptualizes the response to a discrete choice experiment as a two-stage process: first the choice of the attribute processing strategy and second the choice among the offered alternatives conditional on the chosen processing strategy. In an empirical application, he finds that the number of attributes attended to increases as the number of attribute levels declines and the range of an attribute increases. Further, increasing the number of alternatives also increased the number of attributes attended to significantly. Clear evidence was found for choice set simplification through addition of attributes (e.g. different components of travel time). Hensher (2010) further investigates different approaches to attribute processing and incorporates three heuristics analysis: Common-metric attribute aggregation, common-metric parameter transfer and attribute non-attendance. By estimating a number of mixed logit and latent class models, he shows that various heuristics in attribute processing influence WTP estimates substantially. While it is relatively easy to estimate the impact of a given choice heuristic, it is more difficult to investigate which strategy was actually chosen by the respondent. Using supporting questions such as "Which attributes did you not attend to?" are convenient, however, Hensher (2010b) proposes delve deeper into the respondent's psyche and apply a Dempster–Shafer belief function to investigate the role of attribute processing in choice experiments.

A growing body of literature has started to tackle the issue of attribute non-attendance (ANA). In principle, "stated" and "inferred" methods of detecting ANA can be distinguished (Kravchenko, 2014). A series of studies have used stated methods to investigate the effects of attribute non-attendance, either on a choice sequence level (Alemu *et al.*, 2013) or at the level of individual choice tasks (Colombo and Glenk, 2013; Quan *et al.*, 2018). Mariel *et al.* (2012) compare stated and inferred ANA methods in the context of wind farms in Germany and conclude that stated ANA is not always consistent with inferred ANA. They estimate inferred ANA by the method developed by Hess and Hensher (2010). Therefore, Alemu *et al.* (2013, p. 341) ask for reasons why a certain attribute was ignored, including (1) the attribute is not important to me, (2) ignoring the attribute made it easier to choose between the alternatives, (3) attribute levels were unrealistically high/low, (4) I do not think the attribute should be weighed against the others and (5) do not know. Alemu *et al.* (2013) argue that reason one exhibits genuine preferences, while reasons three and four exhibit protest behavior. Ignoring the attribute to make the choice easier was specifically often chosen for attributes with non-market good characteristics.

Colombo and Glenk (2013) distinguish between attribute non-attendance and alternative non-attendance in the context of agricultural subsidies of the European Common Agricultural Policy. Based on stated attribute non-attendance after each choice set, they estimate a series of models in which they consider the non-attendance of individual attributes, and find that the benefits of asking debriefing questions after each individual choice would outweigh the additional effort for the respondent. Also, they consider the possibility that an alternative would not be considered at all due to an unacceptable attribute level. They conclude that attribute non-attendance is very common and that the inclusion of the additional information leads to better statistical performance in the estimated models.

Scarpa *et al.* (2009) present an empirical framework to estimate the effect of attribute non-attendance based on a latent class approach and a Bayesian approach. In the latent class approach, they divide their sample into several classes having total attendance to all attributes, total non-attendance (by restricting all parameters to zero) and partial non-attendance (by restricting the parameters of an individual attribute to zero). Further, they investigate non-attendance to combinations of attributes, in particular the interactions of cost and non-monetary attributes. Not accounting for ANA is found to overestimated WTP measures, compared to models where non-attendance, in particular to the combinations with the cost attribute is taken into account. In the Bayesian approach, they account for taste heterogeneity among non-zero taste entities. Findings with regard to WTP were similar to the latent class approach; however, the variance of the estimated parameters was comparably higher. Scarpa *et al.* (2009) conclude that severe attribute non-attendance could be identified from their dataset, and recommend future research to focus on the implications of attribute non-attendance for welfare estimates. In addition, they propose a further investigation into the appropriate supporting questions to better identify attribute non-attending choice strategies.

Hensher *et al.* (2012) provide a probabilistic model to incorporate attribute non-attendance based on a latent class approach. Their basic idea assumes that each respondent is part of one of 2^K classes (with an associated probability) each of which ignores a certain combination of K attributes. This probability is then multiplied with the conditional choice probability of choosing the selected alternative. Similar to Scarpa *et al.* (2009) Hensher *et al.*'s approach has the advantage that it does not rely on stated information about which attributes were not attended to.

Puckett and Hensher (2008) pick up DeShazo and Fermo's (2004) notion of rationally adaptive behavior, "which assumes that decision makers acknowledge that information processing is costly, and hence full attention to the information in a choice task may not be optimal" (Puckett and Hensher, 2008, p. 380). They used two follow-up questions after each

choice set to account for possible adaptively rational behavior, including ignored attributes (either all, or for a specific alternative) and the possibility of adding up attributes along a common dimension (e.g. all cost attributes). Puckett and Hensher (2008) also confirm that incorporating attribute-aggregation and ignored attributes into model estimation leads to very different WTP estimates compared to the standard bounded rational model.

4.3 Order effects
The desire to gain additional information from each respondent, and thereby bring down costs of survey implementation, has led to the inclusion of multiple independent choice sets into a single questionnaire. Standard microeconomic theory conceptualizes these choices to be driven by

(1) All respondents truthfully answer the questions being asked and,

(2) True preferences are stable over the course of a sequence of questions (McNair *et al.*, 2011, p. 556).

With regard to the choice set order, this means that respondents will not be influenced by the order in which choice sets are presented. A number of studies have contested this idea and investigated so-called *order effects* (Bateman *et al.*, 2008; Carson and Groves, 2007; Day *et al.*, 2012; Day and Pinto Prades, 2010; McNair *et al.*, 2011; Scheufele and Bennett, 2012; Vossler *et al.*, 2012). Day *et al.* (2012) divide order effects into *position-dependent* and *precedent-dependent* order effects. Position-dependent order effects influence the respondent's choice because of their position within a series of choice sets. This includes, for example *institutional learning* (Scheufele and Bennett, 2012): a respondent might have had undeveloped preferences for the good in question, which he learns of during the task of going through the choice sets. One way of tackling this would be the inclusion of "warm-up" choice sets preceding the series of choice sets (Carson *et al.*, 1994). However, Meyerhoff and Glenk (2013) point out that this strategy might actually induce starting point bias. Using a split sample approach, they compared samples with and without warm-up choice sets. They found significant differences in WTP between the samples and conclude that including warm-up choice sets "might do more harm than good" (Meyerhoff and Glenk, 2013, p. 25).

Becoming more familiar with the choice context might also induce strategic learning, where a respondent alters his choice behavior for to make a specific strategic goal more likely, without a change in their preferences. Offering a second choice set might also alert the respondent toward the possibility of the price being uncertain or open to negotiation (Carson and Groves, 2007). Being associated with the uncertain variance in future income, WTP might decline for subsequent choice sets. Finally, respondents might become fatigued by the number of choice sets and fail to carefully consider all attributes toward the end of the experiment. They might devolve into satisficing strategies to reduce their cognitive load or show bias for status quo (Day *et al.*, 2012; Giampietri *et al.*, 2016).

In precedent-dependent strategies, the attributes of the previous choice sets affect current choices. For example, to achieve low-cost provision of a public good, a respondent might systematically reject all alternatives being offered at a higher than the lowest price observed thus far. McNair *et al.* (2011, p. 556) name a similar concept, *cost-driven value learning:* an alternative is more (less) likely to be chosen if its cost level is low (high) compared to the levels in the previous choice task(s). Undefined preferences might also induce starting point effects, where respondents compare the attributes of the current choice set to the first-choice set (Day *et al.*, 2012; Ladenburg and Olsen, 2008).

Only few studies have tried to isolate various types of order effects in DCEs empirically. Research on order effects was pioneered by studies examining the double-bounded contingent valuation format (e.g. Hanemann *et al.* (1991) and Cameron and Quiggin (1994)),

where the evidence suggested that WTP estimates from the first- and second-valuation question were not drawn from the same distribution. These findings critically influence the credibility of WTP estimates of DCEs, where a series of seemingly independent choice questions are asked. Empirical evidence, however, suggests that order effects in DCEs exist. For example, Scheufele and Bennet (2012) found that responses significantly depend on previous levels of the cost attribute. In particular, if the level is the highest in the series observed so far, respondents are less likely to choose this alternative in a binary choice setting with status quo and one alternative. Having the minimum cost in the series so far, however, did not show a significant improvement in choice probability. Similar results were found by Day and Pinto Prades (2010) and Zhou et al. (2017). Further, Scheufele and Bennett (2012) observed a significant decline of WTP as the respondent progressed through the series of choice sets. This led them to reject their hypothesis of respondents making stable choice decisions across the sequence of choice sets. McNair et al. (2011) examined if the knowledge of further choice sets influenced the choice in the initial choice set, compared to a part of the sampled respondents who were only provided with a single choice set per questionnaire. While they could reject this hypothesis, they also confirm a significantly lower WTP in subsequent choice sets. Their findings suggest the existence of cost-driven value learning and a possible combination of weak strategic misrepresentation and reference point revision. Further research should focus on the influences of socio-economic influences of observed order effects (McNair et al., 2011).

Day et al. (2012) test a whole series of order effects in their study of preferences for tap water quality improvements. First, they find that the probability of choosing the status quo, regardless of the alternative's attribute levels, is influenced by whether choice sets are revealed sequentially (STP) or in advanced disclosure (ADV) formats. With regard to the presentation mode, they find strong evidence for position dependence in the STP, but not in the ADV mode. As position dependence of STP converges to the level of ADV toward the end, Day et al. argue that institutional learning (in contrast to preference learning) is likely to occur in STP mode. Toward the end of the experiment, the status quo option was significantly more often chosen in STP than in the ADV mode. This can be explained by the theory of loss of credibility as more different combinations at different costs are observed. However, fatigue and the aforementioned income uncertainty hypothesis might also be a reasonable explanation. Day et al. also find evidence of precedent-dependence; however, this is observed in both treatment groups. They use the water quality improvement per cost unit as preference-weighted "deal" measure, and calculate a vector of deal-measures including first task, directly previous choice task, best task and worst task thus far in the series of tasks. While the first deal and the best deal so far significantly shaped preferences of the current choice in both ADV and STP treatment, the worst choice was only significant in the STP treatment, and the directly previous choice was not significant at all. Two explanations for this asymmetry are offered, including either a more cautious perspective of respondents in the STP treatment or strategic misrepresentation of their preferences in the ADV treatment.

5. Choice set design and attribute selection
The selection of attributes through qualitative processes is one of the key issues when designing a choice experiment. In the words of Louviere et al. (2000), "We cannot overemphasize how important it is to conduct this kind of qualitative, exploratory work to guide subsequent phases of the SP study". Despite this recommendation, the documentation of the process of attribute selection in the literature has been sparse (Elgart et al., 2012). However, recently a number of studies in the medical field have been published describing in detail their method of selecting the attributes and levels in a choice experiment (Abiiro et al., 2014; Coast et al., 2012; Kløjgaard et al., 2012; Michaels-Igbokwe et al., 2014). Coast et al. (2012, p. 731)

criticize that with regard to attribute selection, studies hardly report information on sampling, recording, transcription or analytical methods in empirical qualitative studies, or information on search terms, inclusion and exclusion criteria and data extraction methods in literature reviews. Important characteristics of attributes include (1) importance to the respondent for the decision, (2) sufficient distance to the latent construct investigated in the choice experiment (e.g. overall utility should not be an attribute if the researcher is trying to estimate a utility function), (3) single attributes should not have such a large impact that a large number of respondents make no errors in decision making and (4) attributes should not be intrinsic to a person's personality (Coast *et al.*, 2012, p. 734). They review eight papers in the health economics context and make a strong argument for qualitative research approaches in attribute development. They argue that qualitative methods enable the researcher to develop richer and more nuanced attributes than when just taking attributes from the literature or tailored toward a particular policy question. Also, qualitative research could help in refining the language in questionnaires so respondents understand the meaning desired by the researchers. However, they also report some challenges in applying qualitative methods, such as the opportunity costs in generating qualitative research skills, and the reluctance of experienced qualitative researchers to boil down complex relationships into simple and easily comprehensible attributes. Finally, Coast *et al.* (2012) provide a guideline to how attribute development should be reported, including a rationale for the method to develop the attributes, type of sampling and information on how interviews were conducted, details of the analysis, a description of the results and which attributes were problematic and how they were changed or removed from the experiment.

Abiiro *et al.* (2014) pick up Coast *et al.*'s (2012) suggestions and describe in detail their approach to attribute design in a choice experiment in the context of micro-finance health insurance in Malawi. As most studies, they start with a literature review and extract the most important attributes. These attribute lists are used to develop a semi-structured guide for a qualitative study including members of the target population. Abiiro *et al.* (2014) stress that a literature review alone may not capture important attributes specific to the local population. Therefore, community members were led through focus group discussions using open-ended questions, and interviews were conducted with key informants from the health industry. The attributes and their levels were then directly extracted from transcripts using qualitative data analysis, and further narrowed down through additional expert interviews. Criteria for dropping attributes were overlapping with other attributes, attributes where a clear preference was already visible from the focus group discussions to avoid dominance and attributes that had been identified of being less import. All the dropped attributes were fixed to a standard level described in the introduction of the choice experiment. Finally, Abiiro *et al.* (2014) conclude that their qualitative framework could be complemented by basic quantitative methods such as best-worst scaling and nominal group ranking techniques.

Similarly, Michaels-Igbokwe *et al.* (2014) use a mixture of focus group discussions and key informant interviews to select attributes in their choice experiment on health services for young people in Malawi. However, they first engage in a decision mapping process, where they first explore the possible motivations for young people to require access to health services (e.g. those who had used the service before and those who had not – then delve deeper into the reasons why some young adults had not used them). This allowed them to structure their experiment accordingly.

Kløjgaard *et al.* (2012) report that their experience using qualitative processes in designing a choice experiment in the context of degenerative disc diseases of the spine. A literature review to better understand the decision-making context revealed the most relevant patient groups, and also surfaced some instances where patients would regret their decisions to have surgery ex-post. In addition, Kløjgaard *et al.* (2012) conducted three days of observational

field work in a spine surgical treatment ward, observing the patients' questions, thoughts, and motivations and conducting interviews with doctors. After these phases, a first list of attributes and levels was proposed and a preliminary questionnaire was developed. In-depth interviews were conducted with two doctors and three patients, discussing the chosen attributes, which led to the revision of some attributes and levels. In particular, patients were asked whether the attributes should be included, whether some of them were connected, whether the formulation was understandable and if they felt any dominance among the attributes. Next, level formulation and range were discussed, as well as whether a labeled or unlabeled design should be used. In particular, it was found that a labeled (here: surgical vs non-surgical treatment) design might bias a respondent with pre-formed preferences toward the preferred label, without considering the rest of the attributes. Finally, the framing and the total design and layout were discussed with regard to their comprehensibility, length and complexity. Based on these interviews, the attribute list was revised again before conducting a quantitative pilot test.

All of the studies discussed above conclude that qualitative processes are important in attribute design, in particular when shaping the experiment to a certain local context. However, they also point out some difficulties in conducting qualitative research, including the effort and required research skills and difficulty to reduce the wealth of information obtained into a few simple attributes. It would make sense to extend the experience from health economics studies to other fields, such as environmental valuation studies in different cultural contexts.

6. Experimental design

The experimental design is at the core of the choice experiment. It assures the unbiased/uncorrelated distribution of attributes and levels among choice sets, and therefore significantly impacts the consistency and efficiency of estimated parameters. The researcher has to choose which design from the many available (e.g. randomly drawn designs, orthogonal main effects designs, various efficient designs or full factorial designs) for the specific research task. This choice critically depends on the performance of these designs with respect to estimating utility functions. While the classic indicators of design quality were orthogonality (i.e. no attributes correlated with each other) and attribute balance (each attribute occurs exactly the same number of times throughout the design), recent developments have relaxed the orthogonality condition in favor of an efficiency measure. Orthogonal designs for the specific task can usually be found in catalogs, and choice sets can be constructed by randomly pairing alternatives with each other. Other alternatives include cycling through the attributes with each alternative or by the mix-and-match method described by Johnson *et al.* (2007).

Efficient designs minimize some form of error term, in most cases the D-error (Huber and Zwerina, 1996). The D-error is calculated through the asymptotic covariance matrix of the parameter estimates (Street and Burgess, 2007). In linear models, the asymptotic covariance matrix is approximated by the inverse of the second derivatives of the estimated function. In nonlinear models, such as the multinomial logit model and its generalizations, this covariance matrix is calculated through the matrix of second derivatives of the log-likelihood function. In particular, the value being minimized is the determinant of the inverse of the negative expected value of the matrix of second derivatives of the log likelihood function:

$$\Omega_T = -E\left(\frac{\partial^2 L_T(\beta, \lambda)}{\partial(\beta, \lambda)\partial(\beta, \lambda)^{'}}\right) \tag{4}$$

and

$$D = |\Omega_T^{-1}| \qquad (5)$$

This expression is fully general and can be applied to a wide range of models, e.g. the multinomial logit, nested logit, random parameters logit, etc. While most designs used in environmental valuation studies have assumed a linear utility function, recent research has shown an increase in the use of non-linear functions to account for the non-linearity of the multinomial logit model or other models used in the analysis.

Sándor and Wedel (2001) demonstrate their use of a Bayesian design procedure that assumes a prior distribution of likely parameters. In particular, they use marketing manager's prior beliefs about the market share of some product to construct the prior parameter distributions, and then construct the experimental design by minimizing the Bayesian D_B-error (i.e., the expectation of the former mentioned D-error over the prior distribution of the parameter values). They use a multinomial logit model to investigate and minimize the D_B-error.

Bliemer *et al.* (2009) show how to generate efficient designs for analysis with nested logit models and find a significant increase in efficiency compared to standard orthogonal designs. However, this requires a correct specification of the priors. By using Bayesian priors, the design can be made less sensitive to incorrect prior specifications. Also, the efficiency of this particular design is sensitive to misspecifications of the nesting structure. Bliemer and Rose (2010) develop a design strategy for experiments analyzed with the random parameters logit model that allows for correlations across observations. Similar to the nested logit design, the random parameters logit designs are also sensitive toward misspecification of prior distributions of the parameters. Several researchers (Bliemer and Rose, 2010; Ferrini and Scarpa, 2007) suggest to review the literature, or to conduct a pilot study using an orthogonal design, in order to find fitting the prior parameters or parameter distributions.

A number of studies have performed Monte Carlo analysis with regard to this issue (Carlsson and Martinsson, 2003; Ferrini and Scarpa, 2007; Gao *et al.*, 2010; Lusk and Norwood, 2005). In particular, Ferrini and Scarpa (2007) test whether designs using *a priori* information (including Bayesian designs with weakly informative and informative prior) perform better (in terms of deviations from true parameters) than standard design approaches (a shifted fractional factorial orthogonal design). They found that the Bayesian design, under "good" prior information, is robust to model miss-specification if the sample size is large enough, even more than the designs without prior information. Also, in general, their shifted orthogonal design was superior to the D-optimal design with prior information. Gao *et al.* (2010) extended the study by incorporating the attribute information load into their Monte Carlo experiment. Using one continuous and one non-continuous utility function, they estimated the parameters obtained from different design strategies (randomly drawn, orthogonal main effects, minimal D-optimal and random pairings from a full factorial) to find out which design produces the most efficient WTP measures.

Overall, these studies have one common finding: Increasing the sample size significantly improves the WTP measures toward their true values (Carlsson and Martinsson, 2003; Ferrini and Scarpa, 2007; Gao *et al.*, 2010; Lusk and Norwood, 2005). Also, while Lusk and Norwood (2005) find no statistically significant differences between any of the experimental estimates and the true WTP values, they find that designs that include interactions lead to more precise estimates for WTP. Also, according to their findings, larger designs do not necessarily perform better; a main-effects plus two-way interactions orthogonal design containing 243 choice sets did not provide more efficient estimates than a D-efficient two-way interaction design containing 31 choice sets. This has important implications for questionnaire design, as it allows for less complex questionnaires which could lead to more

accurate information. Gao *et al.* (2010) found that WTP measures have a quadratic relationship with the number of attributes in the choice experiment, and recommend a maximum number of three attributes for discrete utility functions. However, they recognize that many aspects in a choice experiment (e.g. statistical efficiency, cognitive burden, budget constraint. . .) are subject to trading off.

Apart from purely technical considerations in experimental design creation, some contextual implications should be considered as well. For example, the researcher has to decide whether to use a labeled or a generic design (Hensher *et al.*, 2005; Johnson *et al.*, 2007). As the name suggests, the labeled design contains information in its label, and therefore requires a different estimation strategy than a purely generic design where all the information is captured by the attributes. For example, a choice experiment using wine might use different labels to describe production methods (e.g. organic, conventional) in each choice set. The other attributes (taste, price, etc.) might then be analyzed with respect to each production method separately.

Also, different approaches toward the choice of attribute levels have been proposed, specifically in transportation research. In particular, Rose *et al.* (2008) suggest the use of pivot designs instead of absolute attribute levels. The basic idea of a pivot design is to let the respondent enter his status quo alternative (e.g. the attributes of his regular transportation to work) and then have the alternatives pivot around this base alternative. While this approach allows for more flexibility, Hess and Rose (2009) list a number of cautions when using data from this type of design. By analyzing data from a transport choice experiment, they found correlations in the error terms across the replications of the reference trips, as well as differences in the variance between the reference trip and the hypothetical trips. Also, their findings suggest asymmetric preference formation around the reference attribute levels.

An important consideration is the cognitive burden that the respondent has to go through when working through the choice tasks. In most cases, the orthogonal or efficient design will still be too large for one respondent to handle. Therefore, designs can be split into multiple blocks, by using blocking algorithms that try to keep the within-block orthogonality maximal (Wheeler, 2011). Kessels *et al.* (2011) introduces a different approach to reduce the cognitive burden by use of partial profiles. They describe a two-stage procedure that generates Bayesian D-optimal designs. In essence, these designs keep some attributes constant over all alternatives, and therefore reduce the cognitive burden on the respondent. They also provide instructions for creating utility-neutral designs, i.e. designs which make the choice probabilities of all alternatives all equal. Kessels *et al.* (2011) conclude that their designs are about 10–20% less efficient as full profile designs, however this drawback might be compensated by the chance of respondents making non-compensatory decisions in designs that are more difficult to evaluate (i.e. not attending to all attributes).

7. Survey mode and sampling

The mode of surveying has been found to influence results of stated choice questionnaires. Commonly available modes include face-to-face (f2f) interviews, mail surveys, telephone interviews and online questionnaires (Champ and Welsh, 2007). As Champ and Welsh (2007) and Dillman and Christian (2005) provide an excellent overview of different survey strategies and their pros and cons, we will only briefly discuss the most important pitfalls that can occur when deciding on a survey mode, and discuss recent findings on differences in responses between different survey modes. Research on differences between survey modes has been conducted on contingent valuations (Macmillan *et al.*, 2002; Maguire, 2009; Marta-Pedroso *et al.*, 2007), but less in choice experiments. The exception is Olsen (2009), who conducted a choice experiment on protecting different types of landscape encroachment in Denmark. They compared response characteristics between an online panel and a random mail survey. While they found only a small difference in response rates between the two survey modes,

they report a significantly larger number of protest bids in the mail sample. Further, while WTP estimates do not differ significantly, they found more homogeneous preferences in the mail vs the online sample. Regarding socio-demographics, the two samples did not differ significantly.

A key issue in the choice of survey mode is the sampling frame, and whether it can be reached better or worse by a specific mode. For example, a population of elderly might be less likely to have Internet connection at home, and therefore might be excluded from online surveys targeted at the general public. Also, there might be systematic demographic or attitudinal differences between people who use online surveys and people who respond to mail questionnaires, depending on the context of the survey (Marta-Pedroso *et al.*, 2007). While researchers from different fields recommend the f2f method (Arrow *et al.*, 1993), issues have been raised with regard to increased social desirability bias and interviewer effects (Ethier *et al.*, 2000; Leggett *et al.*, 2003; Maguire, 2009), resulting in a higher reported WTP than in self-administered modes.

8. Estimation strategy
8.1 Frequentist inference
In principle, one can analyze a choice experiment according to different decision rules. The classic random utility maximization rule (McFadden, 1974), which is described in Section 3, or according to random regret theory (Chorus *et al.*, 2008). In the context of random utility maximization, the multinomial logit and probit model are the most basic models. The probability of choice in the multinomial logit model takes the form

$$P(i) = \frac{\exp(x'\beta)}{\sum_{j=1}^{j} \exp\left(x_j'\beta_j\right)} \tag{6}$$

which can be straightforwardly estimated by maximum likelihood. Some researchers also include individual specific constants into this functional form and therefore estimate a conditional logit model (McFadden, 1974).

However, these models come with the restrictive assumptions of independence of irrelevant alternatives (IIA). To allow for correlation between individual alternatives, the nested logit model can be used to group alternatives according to some criteria. This imposes the assumption on the preference structure that the choice of the individual is first taken between the groups, and then within the groups. For example, a choice set in transportation research might consist of a private car, bus or train. The respondent might first choose among public vs private transport, and then (after choosing public transport) choose between bus and train. This can be modeled by imposing a nested structure on the decision rule.

Several models have been proposed to account for preference heterogeneity among individuals in particular, latent class models (Greene and Hensher, 2003), error components logit (Hensher *et al.*, 2007) and random parameters logit (McFadden and Train, 2000). While multinomial and nested logit models only allow for individual preferences (i.e. marginal utilities) to be fixed and equal across the population (except for interactions with individual-specific characteristics), the random parameters logit model allows the researcher to impose some distribution on the parameters. With this specification, model parameters can be positive, as well as negative or (near) zero for some parts of the population, which adds more realism to the model. A fourth class of models is the latent class model. Similar to the random parameters model, parameters are assumed to vary within the population, however, parameters are discretely distributed. This usually allows to estimate entire parameter sets for a specified number of "latent classes". All of the mentioned estimation methods have been well established in the literature and can be estimated using standard free and open source

software like R (R Core Team, 2014) or Biogeme (Bierlaire, 2003) or commercial (STATA (© StataCorp), Nlogit (© Econometric Software, Inc.), statistical packages (see, e.g. (Train, 2009) for further details on estimation).

If the researcher chooses to use random regret theory as a decision rule, analysis can be conducted similar to the above. Chorus (2012) show that a simple random regret model can be estimated by

$$P(i) = \frac{\exp(-R_i)}{\sum_{j=1}^{j} \exp(-R_j)} \tag{7}$$

where $p(i)$ is the probability of choosing alternative i, and R_i is the regret function (see Section 3). While this method is relatively new, packages already exist to estimate regret functions (e.g. Biogeme, Nlogit).

While the random regret model is not a superior alternative to the random utility model, it allows the dissection of a sample into different decision strategies, and therefore the analysis of the drivers of these strategies (Hess and Stathopoulos, 2013). In the future, it will be interesting to analyze which individuals maximize utility and which minimize their regret. Applications could include comparisons between individuals of different age, gender, socio-cultural background or ethnicity.

8.2 Bayesian inference

Bayesian estimation adds another way to obtain parameter estimates. As Train (2009) points out, Bayesian statistics provide an alternative view on the nature of parameters. In general, parameters are not seen as fixed, but as following a certain distribution. The researcher starts with some initial guess of this parameter distribution $k(\theta)$ and updates his subjective belief of the distribution as more information is obtained. While the asymptotic properties of Bayesian vs maximum likelihood estimation are identical, estimates can differ due to sampling effects. One advantage of Bayesian estimation is that it does not rely on asymptotic assumptions when calculating the variance–covariance matrix of the estimated parameters. However, this comes at the cost of computational intensity, since closed-form solutions of the required distributions are usually not available.

8.3 Endogeneity

Endogeneity in econometrics refers to a correlation of an explanatory variable with the unobserved error term. Louviere *et al.* (2005) report several sources of endogeneity in stated preference method, including attribute non-attendance mentioned above, social interactions between individuals or strategic behavior. Omitted variable bias can occur if respondents infer values of an omitted attribute from included attributes.

Train (2009) describes three methods to deal with endogeneity, the BLP approach, control functions and a maximum likelihood approach. The BLP approach developed by Berry, Levinsohn and Pakes separates the estimated utility function into two parts: a constant across all markets and products, which represents the average utility gained from each product within each market, but is constant across all individuals; this part captures the endogenous error term that is correlated with another explanatory variable (e. g. price). The second part of the utility function captures the preferences of individuals and may include socio-demographic characteristics, as well as an I.I.D error term. The constant term can be estimated in a choice model including a fixed effect for products and markets, and then further be disaggregated using a linear instrumental variable regression, as the error term is still assumed to be correlated with one of the explanatory variables. An application of the

BLP method using social influence variables to correct for endogeneity in transport mode choice is for example Walker *et al.* (2011).

However, in several cases the endogeneity cannot be absorbed by the product-market constant, in particular when the endogeneity occurs at the individual level. An alternative to the BLP method is the control function method (Hausman, 1978; Heckman, 1978; Petrin and Train, 2010), which is set up similarly to simultaneous equation models. The idea in the control function method is to recover the part of the error term that is correlated with an explanatory variable in an instrumental variable regression, and then use the residuals of this equation in the estimation process of the choice model. The control function can be any type of function that describes the conditional mean of the error term in the endogenous choice model. A simple example by Train (2009, p. 335) is set up as follows:

$$U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \varepsilon_{nj}$$
$$y_{nj} = W(z_{nj}, \gamma) + \mu_{nj}$$

(8)

where the utility function U_{nj} depends on the endogenous variable(s) y_{nj}, exogenous variables x_{nj}, and the estimated marginal utilities β_n. The endogenous variable y_{nj} is explained by instruments z_{nj}, parameters γ, and the error term μ_{nj}. The system is estimated in two steps: First, the instrumental variables regression explaining y_{nj} is estimated using for example OLS, and the residuals are recovered. In the second step, the residuals are used to construct the conditional expectation of the error term ε_{nj} in the utility function. If the conditional mean of ε_{nj} is a simple linear function of μ, a parameter for μ can be estimated by simply adding μ to the utility function. Depending on the assumptions on the distribution of μ and ε, e.g. whether the error terms are correlated across alternatives or not, different types of models can be estimated using probit, logit or mixed logit. To incorporate preference heterogeneity, the control function can also be interacted with socio-demographic characteristics (Petrin and Train, 2003).

A similar method to the control function approach is the maximum likelihood approach (Train, 2009), sometimes called Full Information Maximum Likelihood (FIML). Rather than estimating the choice equation in two steps, both equations are estimated in one step. This requires the researcher to specify the joint distribution of μ and ε.

Several studies have applied these methods for dealing with endogeneity, either empirical or by using Monte Carlo simulations. One of the classical examples of endogeneity is missing variable bias. Petrin and Train (2010) apply both the BLP and the control function approach to data from household television services (antenna, cable, cable with added premium, satellite dish). By correcting for endogeneity using the control function approach, they report several parameters switching to the correct signs compared to a model without control function. Petrin and Train point out that the BLP approach is more difficult to implement than the control function approach, as BLP requires a contraction procedure. Also, because it tries to match predicted market shares to true market shares, it is not consistent in the case of sampling error. Overall, their application finds similar estimated parameters and elasticities for both approaches.

Guevara and Ben-Akiva (2010) use both, the two-step control function approach and the FIML approach to investigate the properties of choice models with endogenous variables. Specifically, they show the link between control functions to correct for endogeneity and the use of latent variables (Walker and Ben-Akiva, 2002). Endogeneity often accrues because some qualitative attribute (e.g. comfort) is difficult to measure and therefore not correctly specified in the model. In this case, the latent variable can be explained by an additional equation in a structural equation setting. After estimating an IV regression on the endogenous variable, the residuals are recovered and used as explanatory variable in the latent variable equation, which is integrated out in the final estimation. Alternatively,

the instrumental price equation can be integrated into the latent variable equation directly, and the likelihood function estimated in a single step. Using a series of Monte Carlo experiments, Guevara and Ben-Akiva (2010) found that including a latent variable in their estimation of a control function choice model outperformed both the two-stage control function only and the simultaneous equation control function only models.

Guevara and Polanco (2013) adapt the Multiple Indicator Solution (MIS) method (Wooldridge, 2002) to the use in choice models. Arguing that valid instruments are often difficult to find, Guevara and Polanco (2013) use a system of equations where two indicators are explained by the omitted variable q

$$\begin{aligned} q_1 &= \alpha_0 + \alpha_q q + e_{q_1} \\ q_2 &= \delta_0 + \delta_q q + e_{q_2} \end{aligned} \qquad (9)$$

where q_1 and q_2 are indicators of the omitted variable q, and e_{q1} and e_{q2} are error terms. Under the assumptions that

$$\mathrm{Cov}(q, e_{q_1}) = \mathrm{Cov}(x, e_{q_1}) = \mathrm{Cov}(q, e_{q_2}) = \mathrm{Cov}(x, e_{q_2}) = \mathrm{Cov}(e_{q_1}, e_{q_2}) = 0 \qquad (10)$$

they show that first running a regression of

$$q_1 = \theta_0 + \theta_{q_2} q_2 + \varepsilon, \qquad (11)$$

recovering the residuals $\widehat{\varepsilon}$ and adding them the utility function

$$V_{in} = \beta_0 + \beta_1 x_{1in} + \ldots + \beta_{q_1} q_{1in} + \beta_k x_{kin} + \beta_\varepsilon \widehat{\varepsilon} \qquad (12)$$

can correct for endogeneity. Using Monte Carlo simulations, they find that both, control-functions and MIS, perform similarly well with regard to parameter efficacy and efficiency. However, they find that the MIS method is more robust toward mild violations of the underlying assumptions.

While the control function method leads to consistent ratios between the estimated parameters, the actual estimators are inconsistent (Guevara and Ben-Akiva, 2012). In standard logit choice models, the scale parameter of the extreme value distribution is not identifiable and therefore usually normalized to one. Assuming that the error term of an endogenous choice model $\varepsilon = v + e$, where v is the part correlated with an unobserved variable and e is I.I.D extreme value, Guevara and Ben-Akiva (2012) approximate the joint distribution with an extreme value distribution. They propose a correction for the parameters of the control function model of the size

$$\frac{\mu_{v+e}}{\mu_e} = \frac{\sigma_e}{\sigma} v + e = \frac{\sigma_e}{\sqrt{\sigma_v^2 + \sigma_e^2 + 2\mathrm{cov}(v, e)}} = \frac{\sigma_e}{\sqrt{1 + \sigma_v^2/\sigma_e^2}} \qquad (13)$$

where μ is the scale factor of the extreme value distribution. If the assumption holds that $\sigma_e^2 = \pi^2/3$, leads to a scale factor of

$$\mu_{v+e} = \frac{1}{\sqrt{1 + 3\sigma_v^2/\pi^2}} \qquad (14)$$

Since the ratio between the coefficients is still estimated consistently, Guevara and Ben-Akiva (2012) investigate the effect of omitting a correction of the scale factor on model elasticities and forecasting by Monte Carlo simulation. Their results show that omitting an orthogonal variable affects the scale of the parameters in the logit model, but not the ratios between parameters on a significant level. However, omitting a variable that is correlated with some

other explanatory variable affects the scale, as well as the ratio. Finally, using the two-stage control-function method re-established a consistent estimate of the ratio, but the parameter scale was also affected. With regard to forecasting properties, similar results were found, with the forecast choice probabilities not being affected by the scale issues. Including the residuals from the first stage of the two-stage control function approach in the utility function significantly improved the forecast, while only adjusting for scale performed poorly. In addition, Guevara and Ben-Akiva (2012) apply their ideas to real housing market data. Similar to their simulations, they find that the effect of price on choice is underestimated in a model where quality attributes are not accounted for. Also, they find that other effects (which are correlated with price and quality) are underestimated without the correction for endogeneity. They stress the important policy implications of these findings.

8.4 Demographic variables
Several methods exist to account for preference heterogeneity into discrete choice studies. One way of doing so is to include socio-demographic variables into the choice sets. However, it has to be kept in mind that only the difference between two alternatives counts in the utility function. If the study uses a labeled design, individual-specific intercepts for the alternatives can be estimated. On the other hand, if a generic design is used, socio-demographics have to be interacted with some attribute in order to generate meaningful results (Hensher *et al.*, 2005). This leads to the convenient interpretation of how a certain consumer group likes or dislikes a certain attribute, and adds flexibility to modeling and market share forecasts.

9. Welfare measures
The standard welfare measure when using choice experiments to predict market shares and welfare changes is the compensating variation (CV). In short, the CV is defined as the amount of income that has to be detracted from an individual in order to make him or her as well of as before a price or policy change (Just *et al.*, 2004). For a policy change, the CV can be calculated as

$$V_0\big(p_0, q_0, y\big) = V_1\big(p_1, q_1, y - \text{CV}\big) \tag{15}$$

where the V's are the respective levels of utility, p's are vectors of market goods, q's are vectors of non-market goods and y is income. This can be easily calculated numerically using the goal-seek function of a spreadsheet application. For a marginal change in one attribute, the marginal CV is simply calculated by dividing the parameters $-\beta_{\text{attr}}/\beta_{\text{price}}$, where β_{attr} represents the marginal utility obtained from the attribute, and β_{price} represents the marginal utility of money. The standard errors of the (marginal) CV can be obtained via bootstrapping (Krinsky and Robb, 1986). In random parameters models, the calculation of the distribution of the CV is more difficult and can be obtained via Cholesky factorization of the variance–covariance matrix (see (Hensher *et al.*, 2005) for a short description of the procedure).

A different approach to estimating the standard error of welfare measures is to estimate the model in willingness-to-pay space, instead of preference space (Train and Weeks, 2006). Here, the estimated parameters can be interpreted as marginal willingness-to-pay for the attribute in question right away.

10. Conclusions and further research
Choice experiments have come a long way since their introduction in transportation research. The ecological multifunction of forests has been widely recognized in age of aggravating environmental pollution and biodiversity loss. Elicitation of the forestry non-market values with choice experiments has been intensively studied, such as recreation (Sælen and Ericson, 2013; Juutinenab *et al.*, 2014), carbon sequestration, biodiversity conservation and ecological services (Baranzini, 2012; Cerda, 2006, Cerda *et al.*, 2014).

While the method has improved in nearly all its aspects, a large number of issues still remains. In this review, we have focused on the theoretical and methodological issues that occur in choice experiments and that are continuously identified and discussed in the literature. From the theory point of view, the departure from classical utility theory might be the most important innovation in recent years. It will be interesting to see how this methodological framework is developed further and in what context it can and should be applied in the future.

However, issues that are endogenous to the choice process still remain, no matter which decision rule is used. First of all, the incentive compatibility of hypothetical choice experiments has been challenged by a number of studies, and it will require more research to find if valid and reliable welfare measures can be extracted from such studies. The notion of consequentiality (Vossler et al., 2012) provides an interesting starting point for the development of new elicitation techniques and their application. Further, it puts a stronger focus on the policy implementation, which might often be neglected in environmental valuation studies. On the other hand, correction mechanisms for hypothetical responses in case of overstated willingness-to-pay may be developed.

The body of work related to order effects suggests that severe order effects often exist in choice studies, which should not be neglected. While a number of causes for order effects have been discussed above, the key issue on how to systematically deal with these effects is still largely undetermined. For example, a lower WTP toward the end of a series of choice sets might be caused by institutional learning, by fatigue or by preference learning. This means that either the WTP stated in the first choice sets might be biased upward or the WTP from the last choice sets might be biased downward. Further research is required to identify the "true WTP" from possibly biased responses introduced by the design of the questionnaire.

The order effects issue is also tightly connected to attribute processing. While attribute processing in individual choice sets has been well researched, the question arises if attribute processing strategies change over a series of choice sets. Further, while conceptual models have been developed to deal with attribute non-attendance and other issues, a major issue still is the identification of different processing strategies from the questionnaire. In particular, explicit and implicit methods can be distinguished. In explicit methods, respondents are asked directly which attributes were not attended to. Insights from psychology might help to identify non-attended attributes, and Hensher's (2007) approach to using Dempster–Shafer belief functions could be further developed. Inferred methods, such as Hensher et al.'s (2012) latent class approach, seek to probabilistically separate respondents into different classes who do not attend to one or several attributes, and therefore the contribution of this attribute to utility should be zero. Over all, we see that while there is a large number of conceptual approaches to the modeling of attribute processing strategies, there is still a lack of knowledge of incorporating these approaches into empirical work and deriving measurable conclusions for environmental valuation studies.

On the experimental design frontier, new designs have been developed that incorporate the non-linear nature of choice models into their efficiency measures, in particular the series of designs by Bliemer and Rose (2010); Bliemer et al. (2009) or Street and Burgess (2007). Pivot designs have further increased the flexibility for tailoring experimental designs to the specific (expected) situation. However, this also comes at a cost of more information requirements for picking the right design, and possibly negative consequences if a more sophisticated design is chosen which does not reflect the actual choice situation well. An interesting field of future research might be the optimal design of experiments that incorporate order effects into their optimality measures. This could also be combined with Kessels et al. (2011) approach to partial profiles, to generate designs that reduce the cognitive burden while at the same time reducing the probability of order effects.

Estimation methods are already very advanced, and new theoretical extensions (such as the random regret model) are quickly adopted to complex estimation procedures originally developed for random utility maximization. More and more studies, particularly in the revealed preference field, now consider endogeneity issues in their estimation. Sources of endogeneity in stated preference contexts have been identified, but the literature on estimation in this field is still very sparse, however, problems such as order effects or attribute non-attendance are also related to endogeneity when it comes to estimating utility function. Therefore, more theoretical and empirical work is required on how endogeneity can play a role in stated preference surveys, and what the consequences of endogeneity are when estimating welfare effects.

Overall, even though the method has improved over the recent decades, many issues still remain in practical applications, as well as on a theoretical level. Further research into the handling of these issues is therefore required.

References

Abiiro, G.A., Leppert, G., Mbera, G.B., Robyn, P.J. and Allegri, M.D. (2014), "Developing attributes and attribute-levels for a discrete choice experiment on micro health insurance in rural Malawi", *BMC Health Services Research*, Vol. 14, p. 235, doi: 10.1186/1472-6963-14-235.

Adamowicz, W., Boxall, P., Williams, M. and Louviere, J. (1998), "Stated preference approaches for measuring passive use values: choice experiments and contingent valuation", *American Journal of Agricultural Economics*, Vol. 80, pp. 64-75, doi: 10.2307/3180269.

Alemu, M.H., Mørkbak, M.R., Olsen, S.B. and Jensen, C.L. (2013), "Attending to the reasons for attribute non-attendance in choice experiments", *Environmental and Resource Economics*, Vol. 54, pp. 333-359, doi: 10.1007/s10640-012-9597-8.

Arrow, K., Solow, R. and others (1993), *Report of the NOAA Panel on Contingent Valuation*, National Oceanic and Atmospheric Administration, Washington, DC.

Balcombe, K. and Fraser, I. (2011), "A general treatment of 'don't know' responses from choice experiments", *European Review of Agricultural Economics*. doi: 10.1093/erae/jbr010.

Baranzini, A., *et al.* (2012), "Tropical forest conservation: attitudes and preferences", *Forest Policy and Economics*, Vol. 12 No. 5, pp. 370-376.

Bateman, I.J., Carson, R.T., Day, B., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R. and Wang, P. (2008), "Choice set awareness and ordering effects in discrete choice experiments in discrete choice experiments (No. 08-01)", CSERGE working paper EDM.

Bell, D.E. (1982), "Regret in decision making under uncertainty", *Operations Research*, Vol. 30, pp. 961-981.

Ben-Akiva, M. and Swait, J. (1986), "The akaike likelihood ratio index", *Transportation Science*, Vol. 20, pp. 133-136, doi: 10.1287/trsc.20.2.133.

Bierlaire, M. (2003), "BIOGEME: a free package for the estimation of discrete choice models", *Presented at the Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

Bliemer, M.C.J. and Rose, J.M. (2010), "Construction of experimental designs for mixed logit models allowing for correlation across choice observations", *Transportation Research Part B: Methodological, Methodological Advancements in Constructing Designs and Understanding Respondent Behaviour Related to Stated Preference Experiments*, Vol. 44, pp. 720-734, doi: 10.1016/j.trb.2009.12.004.

Bliemer, M.C.J., Rose, J.M. and Hensher, D.A. (2009), "Efficient stated choice experiments for estimating nested logit models", *Transportation Research Part B: Methodological*, Vol. 43, pp. 19-35, doi: 10.1016/j.trb.2008.05.008.

Boeri, M., Longo, A. and Scarpa, R. (2013), "The regret of not modelling regret in choice experiments: a Monte Carlo investigation", *Presented at the International Choice Modelling Conference*, Sidney, Australia.

Boeri, M., Scarpa, R. and Chorus, C.G. (2014), "Stated choices and benefit estimates in the context of traffic calming schemes: utility maximization, regret minimization, or both?", *Transportation Research Part A: Policy and Practice*, Vol. 61, pp. 121-135, doi: 10.1016/j.tra.2014.01.003.

Boxall, P.C., Adamowicz, W.L., Swait, J., Williams, M. and Louviere, J. (1996), "A comparison of stated preference methods for environmental valuation", *Ecological Economics*, Vol. 18, pp. 243-253, doi: 10.1016/0921-8009(96)00039-0.

Boxall, P., Adamowicz, W.L. and Moon, A. (2009), "Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement*", *Australian Journal of Agricultural and Resource Economics*, Vol. 53, pp. 503-519, doi: 10.1111/j.1467-8489.2009.00469.x.

Brownstone, D. and Small, K.A. (2005), "Valuing time and reliability: assessing the evidence from road pricing demonstrations", *Transportation Research Part A: Policy and Practice*, Vol. 39, pp. 279-293, doi: 10.1016/j.tra.2004.11.001.

Cameron, T.A. and Quiggin, J. (1994), "Estimation using contingent valuation data from a 'dichotomous choice with follow-up' questionnaire", *Journal of Environmental Economics and Management*, Vol. 27, pp. 218-234, doi: 10.1006/jeem.1994.1035.

Carlsson, F. and Martinsson, P. (2003), "Design techniques for stated preference methods in health economics", *Health Economics*, Vol. 12, pp. 281-294, doi: 10.1002/hec.729.

Carson, R.T. and Groves, T. (2007), "Incentive and informational properties of preference questions", *Environmental and Resource Economics*, Vol. 37, pp. 181-210, doi: 10.1007/s10640-007-9124-5.

Carson, R.T., Louviere, J.J., Anderson, D.A., Arabie, P., Bunch, D.S., Hensher, D.A., Johnson, R.M., Kuhfeld, W.F., Steinberg, D., Swait, J., Timmermans, H. and Wiley, J.B. (1994), "Experimental analysis of choice", *Marketing Letters*, Vol. 5, pp. 351-367, doi: 10.1007/BF00999210.

Cerda, C. (2006), *Valuing Biological Diversity in Navarino Island, Cape Horn Archipelago, Chile – a Choice Experiment Approach*, PhD Dissertation, University of Goettingen, Göttingen.

Cerda, C., Barkman, J. and Marggraf, R. (2014), "Non-market economic valuation of the benefits provided by temperate ecosystems at the extreme south of the Americas", *Regional Environmental Change*, Vol. 14, pp. 1517-1531.

Champ, P.A. and Welsh, M.P. (2007), "Survey methodologies for stated-choice studies", in Kanninen, B.J. and Bateman, I.J. (Eds), *Valuing Environmental Amenities Using Stated Choice Studies, the Economics of Non-Market Goods and Resources*, Springer, Netherlands, pp. 21-42.

Chorus, C.G. (2012), *Random Regret-Based Discrete Choice Modeling: A Tutorial*, Springer.

Chorus, C.G., Arentze, T.A. and Timmermans, H.J.P. (2008), "A Random Regret-Minimization model of travel choice", *Transportation Research Part B: Methodological*, Vol. 42, pp. 1-18, doi: 10.1016/j.trb.2007.05.004.

Chorus, C., van Cranenburgh, S. and Dekker, T. (2014), "Random regret minimization for consumer choice modeling: assessment of empirical evidence", *Journal of Business Research*, Vol. 67, pp. 2428-2436, doi: 10.1016/j.jbusres.2014.02.010.

Coast, J., Al-Janabi, H., Sutton, E.J., Horrocks, S.A., Vosper, A.J., Swancutt, D.R. and Flynn, T.N. (2012), "Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations", *Health Economics*, Vol. 21, pp. 730-741, doi: 10.1002/hec.1739.

Colombo, S. and Glenk, K. (2013), "Social preferences for agricultural policy instruments: joint consideration of non-attendance to attributes and to alternatives in modelling discrete choice data", *Journal of Environmental Planning and Management*, Vol. 57, pp. 215-232, doi: 10.1080/09640568.2012.738190.

Day, B. and Pinto Prades, J.-L. (2010), "Ordering anomalies in choice experiments", *Journal of Environmental Economics and Management*, Vol. 59, pp. 271-285, doi: 10.1016/j.jeem.2010.03.001.

Day, B., Bateman, I.J., Carson, R.T., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R. and Wang, P. (2012), "Ordering effects and choice set awareness in repeat-response stated preference studies", *Journal of Environmental Economics and Management*, Vol. 63, pp. 73-91, doi: 10.1016/j.jeem. 2011.09.001.

DeShazo, J.R. and Fermo, G. (2004), *Implications of Rationally-Adaptive Pre-choice Behavior for the Design and Estimation of Choice Models*.

Dillman, D.A. and Christian, L.M. (2005), "Survey mode as a source of instability in responses across surveys", *Field Methods*, Vol. 17, pp. 30-52, doi: 10.1177/1525822X04269550.

Ethier, R.G., Poe, G.L., Schulze, W.D. and Clark, J. (2000), "A comparison of hypothetical phone and mail contingent valuation responses for green-pricing electricity programs", *Land Economics*, Vol. 76, pp. 54-67, doi: 10.2307/3147257.

Ferrini, S. and Scarpa, R. (2007), "Designs with a priori information for nonmarket valuation with choice experiments: a Monte Carlo study", *Journal of Environmental Economics and Management*, Vol. 53, pp. 342-363, doi: 10.1016/j.jeem.2006.10.007.

Fishburn, P.C. (1982), "Nontransitive measurable utility", *Journal of Mathematical Psychology*, Vol. 26, pp. 31-67, doi: 10.1016/0022-2496(82)90034-7.

Freeman, A.M. (2003), *The Measurement of Environmental and Resource Values*, Resource for the Future, Washington D.C.

Gao, Z. and Schroeder, T.C. (2009), "Effects of label information on consumer willingness-to-pay for food attributes", *American Journal of Agricultural Economics*, Vol. 91, pp. 795-809, doi: 10.1111/j.1467-8276.2009.01259.x.

Gao, Z., House, L.O. and Yu, X. (2010), "Using choice experiments to estimate consumer valuation: the role of experimental design and attribute information loads", *Agricultural Economics*, Vol. 41, pp. 555-565, doi: 10.1111/j.1574-0862.2010.00470.x.

Giampietri, E., Koemle, D., Yu, X. and Finco, A. (2016), "Consumers' sense of farmers' markets: tasting sustainability or just purchasing food?", *Sustainability*, Vol. 8 No. 11, p. 1157, doi: 10.3390/su8111157.

Greene, W.H. and Hensher, D.A. (2003), "A latent class model for discrete choice analysis: contrasts with mixed logit", *Transportation Research Part B: Methodological*, Vol. 37, pp. 681-698, doi: 10.1016/S0191-2615(02)00046-2.

Guevara, C.A. and Ben-Akiva, M. (2010), "Addressing endogeneity in discrete choice models: assessing control-function and latent-variable methods", *Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference*, Emerald Group Publishing, p. 353.

Guevara, C.A. and Ben-Akiva, M.E. (2012), "Change of scale and forecasting with the control-function method in logit models", *Transportation Science*, Vol. 46, pp. 425-437, doi: 10.1287/trsc.1110.0404.

Guevara, C.A. and Polanco, D. (2013), "Correcting for endogeneity without instruments in discrete choice models: the multiple indicator solution", *Presented at the International Choice Modeling Conference*, Sidney.

Hanemann, M., Loomis, J. and Kanninen, B. (1991), "Statistical efficiency of double-bounded dichotomous choice contingent valuation", *American Journal of Agricultural Economics*, Vol. 73, pp. 1255-1263, doi: 10.2307/1242453.

Hanley, N., Wright, R.E. and Adamowicz, W. (1998), "Using choice experiments to value the environment", *Environmental and Resource Economics*, Vol. 11, pp. 413-428.

Hanley, N., Mourato, S. and Wright, R.E. (2001), "Choice modelling approaches: a superior alternative for environmental valuation?", *Journal of Economic Surveys*, Vol. 15, pp. 435-62.

Harrison, G.W. (2007), "Making choice studies incentive compatible", in Kanninen, B.J. and Bateman, I.J. (Eds), *Valuing Environmental Amenities Using Stated Choice Studies, the Economics of Non-Market Goods and Resources*, Springer, Netherlands, pp. 67-110.

Hausman, J.A. (1978), "Specification tests in econometrics", *Econometrica*, Vol. 46, p. 1251, doi: 10.2307/1913827.

Heckman, J.J. (1978), "Dummy endogenous variables in a simultaneous equation system", *Econometrica*, Vol. 46, pp. 931-959, doi: 10.2307/1909757.

Hensher, D. (2007), "Attribute processing in choice experiments and implications on willingness to pay", in Kanninen, B.J. and Bateman, I.J. (Eds), *Valuing Environmental Amenities Using Stated Choice Studies, the Economics of Non-Market Goods and Resources*, Springer, Netherlands, pp. 135-137.

Hensher, D.A. (2010a), "Hypothetical bias, choice experiments and willingness to pay", *Transportation Research Part B: Methodological*, Vol. 44, pp. 735-752, doi: 10.1016/j.trb.2009.12.012.

Hensher, D.A. (2010b), "Attribute processing, heuristics and preference construction in choice analysis", *State of Art and State of Practice in Choice Modelling. Presented at the International Choice Modelling Conference*, Emerald Group Publishing Limited, Yorkshire, pp. 35-70.

Hensher, D.A., Rose, J.M. and Greene, W.H. (2005), *Applied Choice Analysis: A Primer*, Cambridge University Press.

Hensher, D.A., Jones, S. and Greene, W.H. (2007), "An error component logit analysis of corporate bankruptcy and insolvency risk in Australia", *Economic Record*, Vol. 83, pp. 86-103, doi: 10.1111/j.1475-4932.2007.00378.x.

Hensher, D., Rose, J. and Greene, W. (2012), "Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design", *Transportation*, Vol. 39, pp. 235-245, doi: 10.1007/s11116-011-9347-8.

Hess, S. and Hensher, D.A. (2010), "Using conditioning on observed choices to retrieve individual-specific attribute processing strategies", *Transportation Research Part B: Methodological, Methodological Advancements in Constructing Designs and Understanding Respondent Behaviour Related to Stated Preference Experiments*, Vol. 44, pp. 781-790, doi: 10.1016/j.trb.2009.12.001.

Hess, S. and Rose, J.M. (2009), "Should reference alternatives in pivot design SC surveys be treated differently?", *Environmental and Resource Economics*, Vol. 42, pp. 297-317, doi: 10.1007/s10640-008-9244-6.

Hess, S. and Stathopoulos, A. (2013), "A mixed random utility — random regret model linking the choice of decision rule to latent character traits", *Journal of Choice Modelling, Issues in Choice Modelling: Selected Papers from the 13th International Conference on Travel Behaviour Research*, Vol. 9, pp. 27-38, doi: 10.1016/j.jocm.2013.12.005.

Hoyos, D. (2010), "The state of the art of environmental valuation with discrete choice experiments", *Ecological Economics*, Vol. 69, pp. 1595-1603, doi: 10.1016/j.ecolecon.2010.04.011.

Huber, J. and Zwerina, K. (1996), "The importance of utility balance in efficient choice designs", *Journal of Marketing Research*, pp. 307-317.

Johnson, F., Kanninen, B., Bingham, M. and Özdemir, S. (2007), "Experimental design for stated-choice studies", in Kanninen, B.J. and Bateman, I.J. (Eds), *Valuing Environmental Amenities Using Stated Choice Studies, the Economics of Non-Market Goods and Resources*, Springer, Netherlands, pp. 159-202.

Just, R.E., Hueth, D.L. and Schmitz, A. (2004), *The Welfare Economics of Public Policy: A Practical Approach to Project and Policy Evaluation*, Edward Elgar.

Juutinenab, A., Koseniusc, A.K. and Ovaskainenc, V. (2014), "Estimating the benefits of recreation-oriented management in state-owned commercial forests in Finland: a choice experiment", *Journal of Forest Economics*, Vol. 20 No. 4, pp. 396-412.

Kanninen, B.J. (2007), *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice*, Springer.

Kessels, R., Jones, B. and Goos, P. (2011), "Bayesian optimal designs for discrete choice experiments with partial profiles", *Journal of Choice Modelling*, Vol. 4, pp. 52-74, doi: 10.1016/S1755-5345(13) 70042-3.

Kløjgaard, M.E., Bech, M. and Søgaard, R. (2012), "Designing a stated choice experiment: the value of a qualitative process", *Journal of Choice Modelling*, Vol. 5, pp. 1-18, doi: 10.1016/S1755-5345(13) 70050-2.

Kravchenko, A. (2014), "Influence of rudimentary attribute non-attendance (ANA) on choice experiment parameter estimates and design efficiency: a Monte Carlo Simulation analysis", *Journal of Choice Modelling, Process Heuristics in Choice Analysis*, Vol. 11, pp. 57-68, doi: 10. 1016/j.jocm.2014.02.002.

Krawczyk, M. (2012), "Testing for hypothetical bias in willingness to support a reforestation program", *Journal of Forest Economics*, Vol. 18, pp. 282-289, doi: 10.1016/j.jfe.2012.07.003.

Krinsky, I. and Robb, A.L. (1986), "On approximating the statistical properties of elasticities", *The Review of Economics and Statistics*, Vol. 68, pp. 715-719, doi: 10.2307/1924536.

Ladenburg, J. and Olsen, S.B. (2008), "Gender-specific starting point bias in choice experiments: evidence from an empirical study", *Journal of Environmental Economics and Management*, Vol. 56, pp. 275-285, doi: 10.1016/j.jeem.2008.01.004.

Lancaster, K.J. (1966), "A new approach to consumer theory", *Journal of Political Economy*, Vol. 74.

Lanz, B. and Provins, A. (2012), *Do Status Quo Choices Reflect Preferences? Evidence from a Discrete Choice Experiment in the Context of Water Utilities' Investment Planning*, CEPE Working Paper Series No. 12-87, CEPE Center for Energy Policy and Economics, ETH Zurich.

Leggett, C.G., Kleckner, N.S., Boyle, K.J., Dufield, J.W. and Mitchell, R.C. (2003), "Social desirability bias in contingent valuation surveys administered through in-person interviews", *Land Economics*, Vol. 79, pp. 561-575, doi: 10.3368/le.79.4.561.

Loomes, G. and Sugden, R. (1982), "Regret theory: an alternative theory of rational choice under uncertainty", *The Economic Journal*, Vol. 92, pp. 805-824, doi: 10.2307/2232669.

Louviere, J.J. and Hensher, D.A. (1982), *Design and Analysis of Simulated Choice or Allocation Experiment in Travel Choice Modelling*, Transportation Research Record.

Louviere, J.J., Hensher, D.A. and Swait, J.D. (2000), *Stated Choice Methods: Analysis and Applications*, Cambridge University Press.

Louviere, J., Train, K., Ben-Akiva, M., Bhat, C., Brownstone, D., Cameron, T.A., Carson, R.T., Deshazo, J.R., Fiebig, D., Greene, W., Hensher, D. and Waldman, D. (2005), "Recent progress on endogeneity in choice modeling", *Marketing Letters*, Vol. 16, pp. 255-265, doi: 10.1007/s11002-005-5890-4.

Lusk, J.L. and Norwood, F.B. (2005), "Effect of experimental design on choice-based conjoint valuation estimates", *American Journal of Agricultural Economics*, Vol. 87, pp. 771-785.

Lusk, J.L. and Schroeder, T.C. (2004), "Are choice experiments incentive compatible? A test with quality differentiated beef steaks", *American Journal of Agricultural Economics*, Vol. 86, pp. 467-482, doi: 10.1111/j.0092-5853.2004.00592.x.

Macmillan, D.C., Philip, L., Hanley, N. and Alvarez-Farizo, B. (2002), "Valuing the non-market benefits of wild goose conservation: a comparison of interview and group based approaches", *Ecological Economics*, Vol. 43, pp. 49-59, doi: 10.1016/S0921-8009(02)00182-9.

Maguire, K.B. (2009), "Does mode matter? A comparison of telephone, mail, and in-person treatments in contingent valuation surveys", *Journal of Environmental Management*, Vol. 90, pp. 3528-3533, doi: 10.1016/j.jenvman.2009.06.005.

Mariel, P., Boeri, M., Meyerhoff, J. and Hoyos, D. (2012), "Dealing with controversial and non-attended attributes in discrete choice experiments", *Presented at the European Association of Environmental and Resource Economists – 19th Annual Conference*, Prague.

Marta-Pedroso, C., Freitas, H. and Domingos, T. (2007), "Testing for the survey mode effect on contingent valuation data quality: a case study of web based versus in-person interviews", *Ecological Economics*, Vol. 62, pp. 388-398, doi: 10.1016/j.ecolecon.2007.02.005.

McFadden, D. (1974), "Conditional logit analysis of qualitative choice behavior", in Zarembka, P. (Ed.), *Frontiers in Econometrics*, Academic Press, pp. 105-142.

McFadden, D. and Train, K. (2000), "Mixed MNL models for discrete response", *Journal of Applied Economics*, Vol. 15, pp. 447-470, doi: 10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1.

McNair, B.J., Bennett, J. and Hensher, D.A. (2011), "A comparison of responses to single and repeated discrete choice questions", *Resource and Energy Economics*, Vol. 33, pp. 554-571, doi: 10.1016/j.reseneeco.2010.12.003.

Meyerhoff, J. and Glenk, K. (2013), *Learning How to Choose – Effects of Instructional Choice Sets in Discrete Choice Experiments*, Working Papers on Management in Environmental Planning, Technische Universität Berlin, Berlin.

Meyerhoff, J. and Liebe, U. (2009), "Status quo effect in choice experiments: empirical evidence on attitudes and choice task complexity", *Land Economics*, Vol. 85, pp. 515-528.

Michaels-Igbokwe, C., Lagarde, M., Cairns, J. and Terris-Prestholt, F. (2014), "Using decision mapping to inform the development of a stated choice survey to elicit youth preferences for sexual and reproductive health and HIV services in rural Malawi", *Social Science and Medicine*, Vol. 105, pp. 93-102, doi: 10.1016/j.socscimed.2014.01.016.

Morey, E.R., Buchanan, T. and Waldman, D.M. (2002), "Estimating the benefits and costs to mountain bikers of changes in trail characteristics, access fees, and site closures: choice experiments and benefits transfer", *Journal of Environmental Management*, Vol. 64, pp. 411-422.

Murphy, J.J., Allen, P.G., Stevens, T.H. and Weatherhead, D. (2005), "A meta-analysis of hypothetical bias in stated preference valuation", *Environmental and Resource Economics*, Vol. 30, pp. 313-325, doi: 10.1007/s10640-004-3332-z.

Olsen, S.B. (2009), "Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods", *Environmental and Resource Economics*, Vol. 44, pp. 591-610, doi: 10.1007/s10640-009-9303-7.

Payne, J.W., Bettman, J.R., Coupey, E. and Johnson, E.J. (1992), "A constructive process view of decision making: multiple strategies in judgment and choice", *Acta Psychologica*, Vol. 80, pp. 107-141, doi: 10.1016/0001-6918(92)90043-D.

Petrin, A. and Train, K. (2003), *Omitted Product Attributes in Discrete Choice Models*, (Working Paper No. 9452), National Bureau of Economic Research.

Petrin, A. and Train, K. (2010), "A control function approach to endogeneity in consumer choice models", *Journal of Marketing Research*, Vol. 47, pp. 3-13, doi: 10.1509/jmkr.47.1.3.

Puckett, S.M. and Hensher, D.A. (2008), "The role of attribute processing strategies in estimating the preferences of road freight stakeholders", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 44, pp. 379-395, doi: 10.1016/j.tre.2007.01.002.

Quan, S., Zeng, Y., Yu, X. and Bao, T. (2018), "WTP for baby milk formula in China: using attribute non-attendance as a priori information to select attributes in choice experiment", *Agribusiness: An International Journal*, Vol. 34 No. 2, pp. 300-320.

R Core Team (2014), *R: A Language and Environment for Statistical Computing*, Vienna, Austria.

Rose, J.M., Bliemer, M.C.J., Hensher, D.A. and Collins, A.T. (2008), "Designing efficient stated choice experiments in the presence of reference alternatives", *Transportation Research Part B: Methodological*, Vol. 42, pp. 395-406, doi: 10.1016/j.trb.2007.09.002.

Sælen, H. and Ericson, T. (2013), "The recreational value of different winter conditions in Oslo forests: a choice experiment", *Journal of Environmental Management*, Vol. 131, pp. 426-434.

Sándor, Z. and Wedel, M. (2001), "Designing conjoint choice experiments using managers' prior beliefs", *Journal of Marketing Research*, Vol. 38, pp. 430-444.

Samuelson, W. and Zeckhauser, R. (1988), "Status quo bias in decision making", *Journal of Risk and Uncertainty*, Vol. 1, pp. 7-59, doi: 10.1007/BF00055564.

Scarpa, R., Gilbride, T.J., Campbell, D. and Hensher, D.A. (2009), "Modelling attribute non-attendance in choice experiments for rural landscape valuation", *European Review of Agricultural Economics*, Vol. 36, pp. 151-174, doi: 10.1093/erae/jbp012.

Scheufele, G. and Bennett, J. (2012), "Response strategies and learning in discrete choice experiments", *Environmental and Resource Economics*, Vol. 52, pp. 435-453, doi: 10.1007/s10640-011-9537-z.

Street, D.J. and Burgess, L. (2007), "The construction of optimal stated choice experiments: theory and methods", *Wiley Series in Probability and Statistics*, Wiley.

Thiene, M., Boeri, M. and Chorus, C.G. (2012), "Random regret minimization: exploration of a new choice model for environmental and resource economics", *Environmental and Resource Economics*, Vol. 51, pp. 413-429, doi: 10.1007/s10640-011-9505-7.

Train, K. (2009), *Discrete Choice Methods with Simulation*, Cambridge University Press.

Train, K. and Weeks, M. (2006), *Discrete Choice Models in Preference Space and Willingness-to Pay Space*.

Von Haefen, R.H., Massey, D.M. and Adamowicz, W.L. (2005), "Serial nonparticipation in repeated discrete choice models", *American Journal of Agricultural Economics*, Vol. 87, pp. 1061-1076, doi: 10.1111/j.1467-8276.2005.00794.x.

Vossler, C.A., Doyon, M. and Rondeau, D. (2012), "Truth in consequentiality: theory and field evidence on discrete choice experiments", *American Economic Journal: Microeconomics*, Vol. 4, pp. 145-171, doi: 10.1257/mic.4.4.145.

Walker, J. and Ben-Akiva, M. (2002), "Generalized random utility model", *Mathematical Social Sciences, Random Utility Theory and Probabilistic Measurement Theory*, Vol. 43, pp. 303-343, doi: 10.1016/S0165-4896(02)00023-9.

Walker, J.L., Ehlers, E., Banerjee, I. and Dugundji, E.R. (2011), "Correcting for endogeneity in behavioral choice models with social influence variables", *Transportation Research Part A: Policy and Practice, Special Issue: Transportation and Social Interactions*, Vol. 45, pp. 362-374, doi: 10.1016/j.tra.2011.01.003.

Wheeler, B. (2011), *AlgDesign: Algorithmic Experimental Design*.

Wooldridge, J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Yu, X., Gao, Z. and Shimokawa, S. (2016), "Consumer preferences for US beef products: a meta-analysis", *Italian Review of Agricultural Economics (REA)*, Vol. 2016 No. 2, pp. 177-195.

Zeelenberg, M. (1999), "The use of crying over spilled milk: a note on the rationality and functionality of regret", *Philosophical Psychology*, Vol. 12, pp. 325-340, doi: 10.1080/095150899105800.

Zhou, S., Yu, X. and Koemle, D. (2017), "Policy choices for air pollution abatement in Beijing: status quo or change", *Singapore Economic Review*, Forthcoming.

Further reading

Koemle, D., Zinngrebe, Y. and Yu, X. (2018), "Highway construction and wildlife populations: evidence from Austria", *Land Use Policy*, Vol. 73, pp. 447-457.

Corresponding author

Xiaohua Yu can be contacted at: xyu@uni-goettingen.de