Environ Resource Econ (2011) 48:129–149
DOI 10.1007/s10640-010-9401-6

Bayesian Conjoint Choice Designs for Measuring Willingness to Pay

Bart Vermeulen · Peter Goos · Riccardo Scarpa · Martina Vandebroek

Accepted: 4 August 2010 / Published online: 12 September 2010

Abstract In this paper, we propose a new criterion for selecting efficient conjoint choice designs when the interest is in quantifying willingness to pay (WTP).The new criterion, which we call the WTP-optimality criterion, is based on the c-optimality criterion which is often used in the optimal experimental design literature. We use a simulation study to evaluate the designs generated using the WTP-optimality criterion and discuss the design of a real-life conjoint experiment from the literature. The results show that the new criterion leads to designs that yield more precise estimates of the WTP than Bayesian D-optimal conjoint choice designs, which are increasingly being seen as the state-of-the-art designs for conjoint choice studies, and to a substantial reduction in the occurrence of unrealistically high WTP estimates.

B. Vermeulen (✉)
Faculty of Business and Economics, Katholieke Universiteit Leuven, Naamsestraat 69, 3000 Leuven, Belgium
e-mail: bart.vermeulen@econ.kuleuven.be

P. Goos
Faculty of Applied Economics & StatUA Center for Statistics, Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen, Belgium
e-mail: peter.goos@ua.ac.be

R. Scarpa
Economics Department, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand
e-mail: rscarpa@waikato.ac.nz

R. Scarpa
School of Agricultural and Resource Economics, University of Western Australia, Perth, Western Australia

M. Vandebroek
Faculty of Business and Economics & Leuven Statistics Research Centre, Katholieke Universiteit Leuven, Naamsestraat 69, 3000 Leuven, Belgium
e-mail: martina.vandebroek@econ.kuleuven.be

P. Goos
Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands

Keywords c-optimal design · Choice experiments · Conditional logit model · D-optimal design · Robust design · Valuation

1 Introduction

Since the early nineties the number of studies using conjoint choice experiments as a tool to estimate the value of attributes of complex goods has vastly increased. Whereas previous studies employing this stated preference method were mostly directed to predict choice behavior and market shares, the increasing emphasis on estimation of implied values of product or service attributes poses new challenges. One such challenge is the development and testing of specific design selection criteria for experiments aimed at estimating the monetary values of attributes and the comparative evaluation with more established criteria. This paper intends to contribute toward this effort.

The objective of conjoint choice experiments is to model respondents' choices as a function of the features of the choice alternatives. For that purpose, the respondents are presented with a series of choice tasks, in each of which they are asked to indicate their favorite alternative. Alternatives are described by means of attributes, each of which has several levels. Because the potential combinations of attribute levels and their allocations in choice tasks are typically many more than can be handled in the course of an interview, experimental design techniques are required to select from the full factorial design a suitable set of choice tasks.

The observed choices are then typically analyzed invoking random utility theory by means of discrete choice models. In valuation studies the estimates of the utility coefficients are often used to calculate marginal rates of substitution (MRS) with respect to the cost coefficient and interpreted as consumers' marginal willingness to pay (WTP) for the attributes. A substantial number of stated preference studies have recently used choice experiments as a tool to obtain WTP estimates. Examples of studies of this kind have been published not only in the conventional fields of application of stated choice experiments, such as in marketing (Sammer and Wüstenhagen 2006), transportation (Hensher and Sullivan 2003), environmental economics (Boxall and Adamowicz 2002 and Adamowicz et al. 1998) and health economics (Ryan 2004), but have also appeared in food (Lusk et al. 2003), livestock (Ruto et al. 2008) and crop research (Kimenju et al. 2005), as well as in cultural (Morey and Rossmann 2003), land (Scarpa et al. 2007) and energy economics (Banfi et al. 2008). In these articles, the use of the conditional logit model has been the dominant approach to data analysis. Therefore, we also focus on the selection of designs for the conditional logit model.

In logit models of discrete choice, the precision of estimates of utility coefficients, and consequently of the marginal WTP, is to a large extent determined by the quality of the data. Thus, the choice of a specific design for any given conjoint choice experiment plays a crucial role. An efficient design maximizes the information in the experiment and in this way guarantees accurate utility coefficient estimates and a powerful statistical inference at a manageable sample size. Creating an efficient conjoint choice design involves selecting the most appropriate choice alternatives and grouping them into choice sets according to a design selection criterion, which is often called an efficiency criterion or an optimality criterion. In this study we focus on an approach that is tailored to the specific problem of estimating functions of utility coefficients, such as the marginal WTP.

The plan of this paper is as follows. In the next section, we discuss the conditional logit model that is typically used to analyze the choices of the respondents and give a brief overview of the conjoint choice design literature to estimate the utility coefficients. In Sect. 3, we first define the marginal WTP and then provide a short overview of the literature on the design of conjoint experiments used for valuation issues. In this section, we present an efficiency criterion for the precise estimation of marginal WTPs and define the corresponding WTP-optimality criterion predicated on weak a-priori information using a Bayesian approach. In Sect. 4, we discuss the results of a simulation study in which designs obtained with different criteria are evaluated in terms of their estimation accuracy for the marginal WTPs. In addition, we examine the designs in terms of their estimation accuracy for the utility coefficients and their predictive performance, which also remain important criteria. Finally, in Sect. 5, we illustrate the performance of the WTP-optimal designs in an example concerning marginal willingness to donate for environmental projects.

2 The Conditional Logit Model

2.1 The Model

Data from a conjoint choice experiment are usually modeled using the widely-known conditional logit model. In the underlying random utility model, the utility of alternative j in choice task k for respondent n is expressed as

$$U_{nkj} = \beta_1 x_{1kj} + \cdots + \beta_M x_{Mkj} + \beta_p p_{kj} + \epsilon_{nkj}. \tag{1}$$

In this model, the first $M + 1$ terms, which we denote by $\mathbf{x}_{kj}^{'}\boldsymbol{\beta}$ in vector notation, form the deterministic component of the utility, and ϵ_{nkj} is the stochastic component representing the response error. The $(M+1)$-dimensional vector $\boldsymbol{\beta}$, which is assumed common for all respondents, contains the utility coefficients of the discrete choice model. These coefficients reflect the importance of the underlying M attributes of the good or service under study and the impact of the price on the utility. The $(M + 1)$-dimensional vector $\mathbf{x}_{kj}$ describes the bundle of these M attributes of alternative j in choice task k and that alternative's price p_{kj}. The response error ϵ_{nkj} captures the unobserved factors influencing the utility experienced by the respondent. In the conditional logit model, the error terms are assumed to be independent and identically Gumbel distributed. The probability that respondent n chooses alternative j in choice task k can then be written as

$$P_{nkj} = \frac{\exp\left(\mathbf{x}_{kj}^{'}\boldsymbol{\beta}\right)}{\sum_{i=1}^{J}\exp\left(\mathbf{x}_{ki}^{'}\boldsymbol{\beta}\right)}, \tag{2}$$

where J is the number of alternatives in choice set k. The conditional logit model assumes that the population under study has homogeneous preferences. This assumption and others underlying the conditional logit model are often criticized for not being realistic (see, e.g., Swait and Louviere 1993; Louviere and Eagle 2006 and Louviere et al. 2002). Nevertheless, this model has proven to be extremely valuable in a large number of recent applications (see, for example, Hearne and Salinas 2002; Brau and Cao 2006; Sammer and Wüstenhagen 2006 and Mtimet and Albisu 2006) and so, remains a basic tool for research in a wide range of areas. The conditional logit model is a corner stone of the panel mixed logit model, so that designs for the conditional logit model are excellent building blocks for constructing designs for the panel mixed logit model which is used in the presence of heterogeneous preferences (see, for example, Yu et al. 2009b and Bliemer and Rose 2010). Remarkably, choice-based conjoint designs based on the conditional logit also perform better than designs obtained by several other frequently used methods. For instance, Bliemer and Rose (2010) show by means of several examples that it is better to use conditional logit designs than orthogonal designs and computationally-intensive cross-sectional mixed logit designs when the interest is in estimating a panel mixed logit model.

2.2 Optimal Designs for Estimating the Utility Parameters

The aim of a conjoint choice experiment is to model how the respondents' choices depend on the attributes of products and services. To achieve this goal in an efficient way, the experiment can be designed such that its information content is maximized. The resulting experimental design is then optimal, at least according to some design selection criterion and a-priori (i.e. pre-data collection) assumptions. Finding an optimal design for a conjoint choice experiment involves selecting the alternatives to be presented to the respondents and arranging these alternatives in choice sets according to some optimality criterion.

As shown in Kessels et al. (2006), the original design principles like orthogonality, (fractional) factorial designs or designs based on level balance, minimum overlap and utility balance (like in Huber and Zwerina 1996 and Kuhfeld et al. 1994) do not necessarily lead to a maximum information content from a statistical perspective. These criteria are appropriate for creating experimental designs for linear models, but not for non-linear models such as the conditional logit model. In order to maximize the statistical information content of a conjoint choice experiment, the design literature distinguishes several criteria to select a designed experiment. The criteria that received most attention in the marketing literature are the D-, A-, G- and V-optimality criteria (see, e.g., Sándor and Wedel 2001 and Kessels et al. 2006). The most widely used of these is the D-optimality criterion.

In general, D-optimal designs minimize the generalized variance of the parameter estimates, as measured by the determinant of the variance-covariance matrix, and thereby the volume of the confidence ellipsoid around $\boldsymbol{\beta}$. As the variance-covariance matrix is inversely proportional to the Fisher information matrix of the parameter estimates, a D-optimal design also maximizes the determinant of the information matrix on the unknown parameters contained within the vector $\boldsymbol{\beta}$. The performance of a design in terms of the D-optimality criterion is expressed by the D-error

$$D\text{-error} = \{\det \mathrm{I}\ (\mathrm{X}, \boldsymbol{\beta})\}^{-\frac{1}{M+1}} = \{\det \mathrm{V}\ (\mathrm{X}, \boldsymbol{\beta})\}^{\frac{1}{M+1}}, \tag{3}$$

where $\mathrm{I}(\mathrm{X}, \boldsymbol{\beta})$ denotes the Fisher information matrix, $\mathrm{V}(\mathrm{X}, \boldsymbol{\beta})$ is the variance-covariance matrix, and the matrix X contains the attribute levels for all the alternatives in the experiment. The conjoint choice design having the smallest D-error is called the D-optimal design.

Because of the nonlinearity of the conditional logit model, the D-error not only depends on the matrix X but also on the unknown model parameters contained within the $\boldsymbol{\beta}$ vector. As a result, prior knowledge of the model parameters is required to develop an optimal conjoint choice design. However, this knowledge is not available at the time decisions need to be made about the experimental design, and hence researchers need to rely on a-priori assumptions. An extensive overview of the approaches and their assumptions to tackle this problem is given in Kessels et al. (2006). The Bayesian approach used in this paper was introduced by Sándor and Wedel (2001): to optimize a design, they assume a prior distribution with one specific value for $\boldsymbol{\beta}$ as a mean and a variance to take into account the uncertainty related to this specific value. This Bayesian approach finally results in a Bayesian optimal design when the expected error over the prior distribution is optimized. A Bayesian D-optimal conjoint choice design is one that minimizes the D-error in Eq. (3) averaged over a prior distribution for the unknown parameter values. The average D-error is then given by:

$$D_b = E_\beta\left[\{\det V(X,\boldsymbol{\beta})\}^{\frac{1}{M+1}}\right] = \int\limits_{\Re^{M+1}} \{\det V(X,\boldsymbol{\beta})\}^{\frac{1}{M+1}}\pi(\boldsymbol{\beta})d\boldsymbol{\beta}, \tag{4}$$

where $\pi(\boldsymbol{\beta})$ represents the prior distribution. The added value of this approach over locally optimal designs obtained by optimizing the design for only one specific value of $\boldsymbol{\beta}$ was not only shown in Sándor and Wedel (2001), but also in Kessels et al. (2006), Ferrini and Scarpa (2007) and Scarpa et al. (2007). Bayesian D-optimal experimental designs are robust in the sense that, unlike locally D-optimal designs, they do guarantee precise estimates and precise predictions over all likely values of the model parameters. The usefulness of the Bayesian approach and its superior performance compared to orthogonal designs is demonstrated in Kessels et al. (2008).

3 Constructing Optimal Designs to Estimate the WTP

In this section, we first define the marginal WTP. Then, we provide a review of the literature on the design of contingent valuation studies and conjoint choice studies to estimate the marginal WTP. Next, we introduce the WTP-optimality criterion we suggest to select designs for conjoint choice studies that aim at WTP estimation.

3.1 The Marginal Willingness to Pay (WTP)

The marginal rate of substitution (MRS) is the rate which measures the willingness of individuals to give up one attribute of a good or service in exchange for another such that the utility of the good or service remains constant. So, the MRS quantifies the trade-off between the two attributes and thus their relative importance. When the trade-off is made with respect to the price of a good or a service, the MRS is called the marginal willingness to pay (WTP). Thus, the marginal WTP for an attribute measures the change in price that compensates for a change in an attribute, while other attributes are held constant.

To estimate the marginal WTP from a conjoint choice experiment, one of the attributes included in the study, and thus in Eq. (1), has to be the price p. Mathematically, the trade-off between an attribute x_m and the price p can be written as

$$\partial U = \beta_m \partial x_m + \beta_p \partial p = 0, \tag{5}$$

from which it follows that

$$\frac{\partial p}{\partial x_m} = -\frac{\beta_m}{\beta_p}. \tag{6}$$

This ratio of the utility coefficients for attribute m and the price p is called the marginal WTP for the attribute m (see, e.g., Hole 2007).

3.2 WTP Estimation in the Design Literature

In contingent valuation experiments, the marginal WTP for a change in an attribute of a product or a service is estimated by asking the respondent whether he/she is prepared to pay a certain amount of money, the bid, for this change. Constructing the most appropriate designs for these types of experiments have been the issue of a number of papers. First, Nyquist (1992) showed that constructing a design for these experiments by a sequential approach using the information of the observations of one group to adapt the bids for the next group results in more accurate WTP estimates. A similar finding is reported in an application using sequential Bayesian design updating in choice experiments by Scarpa et al. (2007). Kanninen (1993) and Alberini (1995) demonstrated that c-optimal designs, which aim at minimizing the average variance of marginal WTP estimates, outperform D-optimal designs and designs constructed by the fiducial method, which aims at providing narrow fiducial intervals for the WTP. Finally, Kanninen (1995) distinguishes two ways to obtain more precise estimates of the marginal WTP. First, an increase of the sample size proportionately decreases the asymptotic variance of the marginal WTP estimates. Second, the choice of an appropriate bid design might also be useful to reduce the bias and the variance of the marginal WTP estimates. Moreover, she states that efficient bid designs avoid extreme bid values.

Despite an increasing number of applications of conjoint experiments for valuation issues, there is almost no literature on the design of conjoint choice experiments to estimate the marginal WTP precisely. The simulation study of Lusk and Norwood (2005) indicates that random designs and orthogonal designs generated including attribute interactions lead to the most precise WTP estimates compared to other designs, among others main effects orthogonal designs. The key feature of the random designs in this study was that all choice sets were randomly picked from the full factorial design for each respondent separately such that each survey for each respondent is unique. However, Carson et al. (2009) are critical about the generalization of these results (see also Lusk and Norwood 2009).

Ferrini and Scarpa (2007) report results on the accuracy of marginal WTP estimates obtained from a shifted design, a locally D-optimal design and a Bayesian D-optimal design. The construction of a shifted design requires a starting design with as many rows as there are choice sets. These rows serve as the first alternatives in the different choice sets of the experiment. The second alternative for each choice set is then obtained by increasing all attribute levels from the starting design by one, except for the highest level of each attribute which is replaced by the lowest level instead. In a similar fashion, the third alternative for each choice set can be generated from the second one. This procedure is repeated until the desired number of alternatives in every choice set is obtained. Ferrini and Scarpa (2007) conclude that substantial improvements in marginal WTP estimation accuracy could be achieved when a Bayesian D-optimal design was used, provided the prior information was sufficiently precise. The gain in precision was largest in cases with small response errors.

Scarpa and Rose (2008) applied the c-optimality criterion to construct a locally optimal conjoint choice design. They concluded that a c-optimal design leads to more accurate WTP estimates than a random design, an orthogonal design, a locally D-optimal design and a locally A-optimal design, which minimizes the average variance of the parameter estimates.

3.3 Bayesian WTP-Optimal Conjoint Choice Designs

In this paper, we assume that the goal of a conjoint choice experiment is to provide an accurate assessment of the marginal WTP for the attributes of a product or service, and we derive conjoint choice designs that have been constructed specifically for that purpose. We refer to these designs as WTP-optimal designs. The key feature of the WTP-optimal designs is that they minimize the sum of the variances of all WTP estimates obtained from the estimated conditional logit model. Although this is not done here, we note that, where appropriate, the criterion can be specialised to a subset of the attributes under consideration, or even to a single one, if necessary. Finally, note that we study only choice designs that have the same choice sets for every respondent. This is most useful when paper and pencil studies are used, when it is expensive to create graphics to visualize the choice options in a choice set, or when a large sample size is costly to achieve for some reason (e.g. short time available, etc.).

The construction of the WTP-optimal designs requires the variances of the WTP estimates to be quantified. As in Kanninen (1993), we approximate these variances using the so-called delta method, which is based on the Taylor series expansion of the WTP-estimates. Using this method, the variance of a given WTP estimate can be approximated by

$$\begin{aligned} var\left(\widehat{\mathrm{WTP}}_m\right) &= var\left(-\frac{\hat{\beta}_m}{\hat{\beta}_p}\right) \\ &= \frac{1}{\hat{\beta}_p^2}\left(var(\hat{\beta}_m) - 2\left(\frac{\hat{\beta}_m}{\hat{\beta}_p}\right)cov(\hat{\beta}_m, \hat{\beta}_p) + \left(\frac{\hat{\beta}_m}{\hat{\beta}_p}\right)^2 var(\hat{\beta}_p)\right). \end{aligned} \quad (7)$$

In general, a marginal WTP can be estimated for each of the attributes in the model. Therefore, for the model in Eq. (1), M different WTP estimates can be computed. As the researcher is often interested in all these M WTP estimates, we suggest seeking a design that minimizes the sum of the variances of all these M estimates. The design selection criterion therefore becomes

$$\text{WTP-error} = \sum_{m=1}^{M} var\left(\widehat{\mathrm{WTP}}_m\right). \quad (8)$$

The WTP-optimal design minimizes this criterion, which is similar to the A-optimality criterion that seeks designs that minimize the sum of the variances of the estimates of the model coefficients (see, for example, Atkinson and Donev 1992). If not all marginal WTP estimates are relevant for the researcher, the criterion can be easily adapted to take only the relevant ones into account. Note that we determined WTP-optimal designs for the complete set of ratios of the form in Eq. (5) such that the WTP-optimality criterion corresponds to the c-optimality criterion, which is defined in Atkinson and Haines (1996) and used in Kanninen (1993), and the variance-minimizing design criterion in Alberini (1995). However, in this paper, the c-optimality criterion is used to develop a design for a conjoint choice experiment.

A design's performance in terms of the WTP-optimality criterion depends on the unknown parameters. Therefore, a Bayesian approach to the construction of WTP-optimal designs is the most natural approach. The Bayesian approach takes into account a priori information about the unknown model parameters, including the uncertainty associated with that a priori knowledge. This is different from the locally c-optimal design approach adopted by Scarpa and Rose (2008), which assumes the parameters are known with certainty.

The Bayesian approach uses a prior distribution $\pi(\boldsymbol{\beta})$ to summarize the available information about the unknown model parameters in $\boldsymbol{\beta}$. The mean of the prior distribution is the researcher's prior guess of the values of the model parameters. The variance of the prior distribution measures the degree of uncertainty associated with that prior guess. A large variance indicates that the researcher is highly uncertain about the prior guess, whereas a small variance indicates that he/she is quite confident about the prior information. The Bayesian WTP-optimal design then is the design that has the best performance in terms of the WTP-optimality criterion averaged over the prior distribution. In other words, the Bayesian WTP-optimal design minimizes the average WTP-error over $\pi(\boldsymbol{\beta})$:

$$\text{WTP}_b\text{-error} = \int \left[\sum_{m=1}^{M} var\left(\widehat{\mathrm{WTP}}_m\right)\right]\pi(\boldsymbol{\beta})d\boldsymbol{\beta}. \quad (9)$$

As there is no analytical expression for the high-dimensional integral in this expression, it has to be approximated numerically by generating a certain number of draws from the prior distribution $\pi(\boldsymbol{\beta})$, and averaging the WTP-error over these draws. The most commonly used types of draws to compute the values of Bayesian design criteria in the conjoint choice design literature are pseudo Monte Carlo draws, but this requires a large number of draws and entails a substantial computational cost. Therefore, we adopted the approach taken by Train (2003), Baiocchi (2005) and Yu et al. (2009a) and used 100 Halton draws from the prior distribution. These Halton draws form a systematic sample from the prior distribution and, compared to the pseudo Monte Carlo sample, give a better approximation of the Bayesian WTP-error at a lower computational cost. The Halton draws were used as an input to the alternating-sample algorithm described in Kessels et al. (2009), which we modified to construct Bayesian WTP-optimal designs.

4 Evaluation of the Bayesian WTP-Optimal Designs

In this section we report the results from a simulation study conducted to evaluate the proposed WTP-optimality criterion. We compare designs constructed using this criterion with several other commonly used designs. First, we describe the computational aspects related to the Bayesian WTP-optimal designs and the benchmark designs included in our study. Next, we discuss the evaluation measures we utilized and report detailed results.

4.1 Designs

We report results for an experiment with twelve choice sets of three alternatives involving two three-level attributes and one two-level attribute, for each of which effects-type coding was used. In addition to these three attributes, the price was also included in the experiment. The price attribute took two levels that were linearly coded as 1 and 2. This implies that the number of utility coefficients, $M + 1$, contained within $\boldsymbol{\beta}$, equals 6.

The prior distribution we used to construct the Bayesian WTP-optimal designs was a normal distribution with mean $[-0.5, 0, -0.5, 0, -0.5, -0.7]$ and variance

$$\begin{pmatrix} 0.5\mathbf{I}_M & \mathbf{0}_{M\times 1} \\ \mathbf{0}_{1\times M} & 0.05 \end{pmatrix},$$

where $\mathbf{I}_M$ is the M-dimensional identity matrix, and $\mathbf{0}_{M\times 1}$ and $\mathbf{0}_{1\times M}$ represent an M-dimensional column vector and row vector, respectively. The first four elements of the mean vector correspond to the utility coefficients associated with the two three-level attributes. The fifth element corresponds to the two-level attribute and the last element is the coefficient of the price attribute. This prior mean expresses the prior belief that higher attribute levels generate a higher utility, except for the price which has a negative impact on the utility. The prior variance of 0.5 for the coefficients of the first three attributes expresses substantial uncertainty about the prior mean (Kessels et al. 2008). The variance of 0.05 for the price coefficient indicates that the sign of that coefficient is known, but not its magnitude.

To evaluate the performance of the Bayesian WTP-optimal design, we compared it with a Bayesian D-optimal design (see Sándor and Wedel 2001; Kessels et al. 2006; Ferrini and Scarpa 2007) and two standard designs obtained using the options ‘complete enumeration’ and ‘balanced overlap’ in Sawtooth Software. The ‘complete enumeration’ option generates a level-balanced design with minimal level overlap within choice sets and maximum orthogonality. We refer to this design as a near-orthogonal design. The ‘balanced overlap’ option

Table 1 Evaluation criteria for the WTP-optimal design and the three benchmark designs

Design type	WTP_b-error	D_b-error	Level overl. (%)	Ut. bal.	Lev. bal.
WTP-optimal	8.136	0.3285	86.11	10.36	not bal.
D-optimal	9.534	0.2730	44.44	9.13	not bal.
Balanced overlap	16.173	0.3516	56.00	8.60	bal.
Near-orthogonal	18.403	0.3270	33.00	8.53	bal.

allows for a moderate attribute level overlap within choice sets. The Bayesian WTP-optimal design as well as the three benchmark designs are displayed in the Appendix.

4.2 Comparison in Terms of WTP_b-Error, D_b-Error, Level Overlap and Utility Balance

The column labeled 'WTP_b-error' in Table 1 displays the WTP_b-errors for the Bayesian WTP-optimal design and the three benchmark designs using the prior distribution utilized to generate the Bayesian WTP-optimal design. As expected, the WTP_b-errors suggest that the Bayesian WTP-optimal design is the most appropriate design to estimate the marginal WTPs accurately, followed by the Bayesian D-optimal design for which the WTP_b-error is almost 20% higher. The errors of the other benchmark designs are more than twice as high as that of the Bayesian WTP-optimal design. Thus, the standard designs perform poorly when it comes to estimating the marginal WTPs: to achieve the same precision from the standard designs as from the optimal designs, twice as many respondents are required. The poor performance of the standard designs is in line with the results reported in Scarpa and Rose (2008).

Moreover, Table 1 shows the D_b-error for the Bayesian WTP-optimal and the three benchmark designs. The D_b-error reflects the average performance of a design in terms of the frequently used D-optimality criterion over the prior distribution and so, is calculated here over the 100 Halton draws used to construct the optimal designs. As the Bayesian D-optimal design by definition minimizes the D_b-error, it has the best performance in terms of this criterion. The WTP-optimal design performs worse, although it is nearly as good as the near-orthogonal design and better than the balanced overlap design in terms of the D_b-error.

Finally, we also studied the level overlap, level balance and utility balance of the designs. Evaluating the WTP-optimal design in terms of these classical design concepts gives an indication of whether or not these features are important to bear in mind when constructing designs to measure the WTPs accurately. The values in the column labeled 'level overlap' are the proportions of columns in the choice sets which exhibit level overlap. We say that a design is level balanced if all levels of the attributes occur equally often in the design. It can be seen that the WTP-optimal design exhibits the highest level overlap and that only the balanced overlap and near-orthogonal designs are level balanced. To measure utility balance, we use the average cumulative entropy as suggested by Swait and Adamowicz (2001) and used in Kessels et al. (2006):

$$-\sum_{k=1}^{K}\int\left(\sum_{j=1}^{J}P_{kj}\ln(P_{kj})\right)\pi(\boldsymbol{\beta})d\boldsymbol{\beta}, \tag{10}$$

where K is the number of choice sets in the design and J is the number of alternatives in each choice set. The higher the cumulative entropy, the closer the alternatives are in terms of utility. A design of the same size and including the same attributes as in our case is maximum utility balanced if the cumulative entropy is 13.18. Table 1 shows that the WTP-optimal design is the most utility balanced of all designs studied, but it is not at all maximum utility balanced. Finally, it can be seen that the D-optimal design is not maximum utility balanced either, which is in line with the results described in Kessels et al. (2006). Also, the balanced overlap and near-orthogonal designs are not utility balanced.

4.3 Simulation Study

4.3.1 Evaluation Criteria

The next evaluation measure we use for the quality of estimation is the expected mean squared error of the WTP estimates. That measure quantifies the difference between the WTP estimates $\widehat{\boldsymbol{W}}(\hat{\boldsymbol{\beta}})$ constructed using the utility coefficient estimates $\hat{\boldsymbol{\beta}}$ with the WTP values $\boldsymbol{W}(\boldsymbol{\beta})$ corresponding to the $\boldsymbol{\beta}$ vector used to simulate data:

$$\text{EMSE}_{\text{WTP}}(\boldsymbol{\beta}) = \int (\widehat{\boldsymbol{W}}(\hat{\boldsymbol{\beta}}) - \boldsymbol{W}(\boldsymbol{\beta}))^{'}(\widehat{\boldsymbol{W}}(\hat{\boldsymbol{\beta}}) - \boldsymbol{W}(\boldsymbol{\beta}))f(\hat{\boldsymbol{\beta}})d\hat{\boldsymbol{\beta}}, \tag{11}$$

with $f(\hat{\boldsymbol{\beta}})$ the distribution of the utility coefficient estimates $\hat{\boldsymbol{\beta}}$. The $\text{EMSE}_{\text{WTP}}(\boldsymbol{\beta})$ value captures the bias and the variability in the marginal WTP estimates. Obviously, a small $\text{EMSE}_{\text{WTP}}(\boldsymbol{\beta})$ value is preferred over a large one. We also calculated the bias between the true and estimated marginal WTP estimates, averaged over $f(\hat{\boldsymbol{\beta}})$:

$$B_{\text{WTP}}(\boldsymbol{\beta}) = \int \left(\hat{\boldsymbol{W}}(\hat{\boldsymbol{\beta}}) - \boldsymbol{W}(\boldsymbol{\beta})\right) f(\hat{\boldsymbol{\beta}})d\hat{\boldsymbol{\beta}}. \tag{12}$$

In the estimation of the marginal WTP values, the price coefficient's estimate plays a very important role as it forms the basis for each individual WTP estimate. A poor estimate of the price coefficient thus results in poor estimates for each individual WTP and high $\text{EMSE}_{\text{WTP}}(\boldsymbol{\beta})$ values. As we shall see below, this problem occurs quite frequently and necessitated us to display the logarithm of the $\text{EMSE}_{\text{WTP}}(\boldsymbol{\beta})$ values. The problem of unrealistic marginal WTP estimates has already been described by Sonnier et al. (2007) and Scarpa et al. (2008), among others.

We also examined the accuracy of the utility coefficient estimates $\hat{\boldsymbol{\beta}}$ themselves. For that purpose, we used the expected mean squared error

$$\text{EMSE}_{\beta}(\boldsymbol{\beta}) = \int (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{'}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})f(\hat{\boldsymbol{\beta}})d\hat{\boldsymbol{\beta}}. \tag{13}$$

A small $\text{EMSE}_{\beta}(\boldsymbol{\beta})$ value is desirable.

Finally, we also calculated the prediction performance of the designs. Using the coefficient estimates, the choice probabilities for each alternative in the twelve choice sets of the design used to simulate the data were computed. Comparing these predicted probabilities with the probabilities based on the 'true' utility coefficients (used to simulate the data) allowed us to evaluate the predictive performance of the designs. We quantified the prediction error using the expected mean squared error

$$\text{EMSE}_{p}(\boldsymbol{\beta}) = \int (\boldsymbol{p}(\hat{\boldsymbol{\beta}}) - \boldsymbol{p}(\boldsymbol{\beta}))^{'}(\boldsymbol{p}(\hat{\boldsymbol{\beta}}) - \boldsymbol{p}(\boldsymbol{\beta}))f(\hat{\boldsymbol{\beta}})d\hat{\boldsymbol{\beta}}, \tag{14}$$

where $\boldsymbol{p}(\hat{\boldsymbol{\beta}})$ and $\boldsymbol{p}(\boldsymbol{\beta})$ are vectors containing the predicted and the 'true' choice probabilities, respectively, for each of the three alternatives in the twelve choice sets. Obviously, small $\text{EMSE}_p(\boldsymbol{\beta})$ values are preferred over large ones.

As different true values of $\boldsymbol{\beta}$ lead to different values of the evaluation measures, we computed $\text{EMSE}_{\text{WTP}}(\boldsymbol{\beta})$, $B_{\text{WTP}}(\boldsymbol{\beta})$, $\text{EMSE}_{\beta}(\boldsymbol{\beta})$ and $\text{EMSE}_p(\boldsymbol{\beta})$ values for 75 different values of $\boldsymbol{\beta}$. For each of the 75 $\boldsymbol{\beta}$ values, we simulated 1000 data sets assuming that there are 75 respondents each time.

4.3.2 Design Performance Under Correct Priors

First we computed the evaluation measures for each of 75 $\boldsymbol{\beta}$ vectors randomly drawn from the prior distribution used to construct the Bayesian WTP-optimal design. As a result, the evaluation measures discussed in this section are representative for a situation in which the prior information about the utility coefficients is reasonably correct.

Figure 1 shows box plots of the logarithms of the 75 average EMSE_{WTP} values for the Bayesian WTP-optimal design, the Bayesian D-optimal design and the two standard designs. It is clear that the Bayesian WTP-optimal design is the most reliable one since it does not only produce the smallest average EMSE_{WTP} value but it also exhibits the smallest spread in EMSE_{WTP} values. Because of the logarithmic scale, the Bayesian WTP-optimal design appears only marginally better than the D-optimal design. However, the raw EMSE_{WTP} values of the two designs are more than different enough to conclude that there is a practical difference between the two designs. This can be clearly seen from Table 2, where we displayed the average EMSE_{WTP} values, their minima and maxima, and the number of outlying EMSE_{WTP} values obtained from using the four designs in our study. The results also show that the difference between the two Bayesian optimal designs, on the one hand, and the standard designs, on the other hand, is even larger. Note that we report two versions of the average, minimum and maximum EMSE_{WTP} values. For each of these statistics, one value was computed based on estimates from all the simulated datasets, whereas the other value was computed after removing the outlying WTP estimates. Even after excluding the outliers for each design option, the Bayesian WTP-optimal design still results in substantially more accurate marginal WTP estimates than the benchmark designs. In this case, the

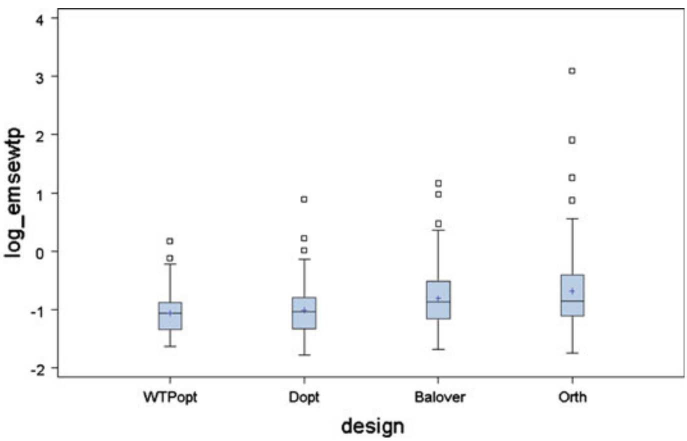

Fig. 1 Log (EMSE_{WTP}) values for the different designs assuming a correct prior distribution

Table 2 Summary statistics of $EMSE_{WTP}$ values with and without outliers over 75 parameter sets $\boldsymbol{\beta}$

Simulation statistics	WTP-opt.	D-opt.	Bal. Overl.	Near-Orth.
Average $EMSE_{WTP}$	0.112 (0.139)	0.125 (0.262)	0.198 (0.579)	0.223 (18.268)
Minimum $EMSE_{WTP}$	0.002 (0.002)	0.002 (0.002)	0.003 (0.003)	0.003 (0.003)
Maximum $EMSE_{WTP}$	0.745 (4.347)	0.895 (90.786)	1.440 (111.700)	1.736 (12893.140)
Average # outliers	9.4	14.4	18.3	24.1

Values obtained with outliers are given in parentheses

average $EMSE_{WTP}$ value when using a D-optimal design is about 10% higher than the average $EMSE_{WTP}$ value when using a WTP-optimal design. The near-orthogonal design exhibits the worst performance, yielding an average $EMSE_{WTP}$ value which is twice as large as that produced by the WTP-optimal design.

A striking result is that the number of outlying $EMSE_{WTP}$ values is much larger for the two benchmark designs than for the Bayesian optimal designs, and that the Bayesian WTP-optimal design produces substantially fewer outliers than the Bayesian D-optimal design. To determine whether or not a marginal WTP estimate was outlying, we compared it with $Q_3 + 6 \cdot IQR$, where Q_3 is the third quartile of the $EMSE_{WTP}$ values and IQR is the interquartile range. The last row of Table 2 shows the average number of outliers among the marginal WTP estimates over all $\boldsymbol{\beta}$ vectors we generated. The implication of the small number of outliers for the WTP-optimal design is that, unlike the other design options, the Bayesian WTP-optimality criterion seems to guarantee that the WTP estimates are seldom completely wrong. This is completely different for the 'balanced overlap' and the nearly orthogonal design.

As the $EMSE_{WTP}$ values summarize the bias and the variance of the marginal WTP estimates, we also studied the bias of the WTP estimates separately. Figure 2 displays the average bias for the 75 $\boldsymbol{\beta}$ values for the marginal WTP for the second level of the first attribute. Similar results hold for the other WTPs. It can be seen that the box plot of the WTP-optimal design has the smallest box, the shortest whiskers and the fewest outlying observations of all designs studied. This indicates that the WTP-optimal design leads to the most accurate estimates of the marginal WTP for that attribute level.

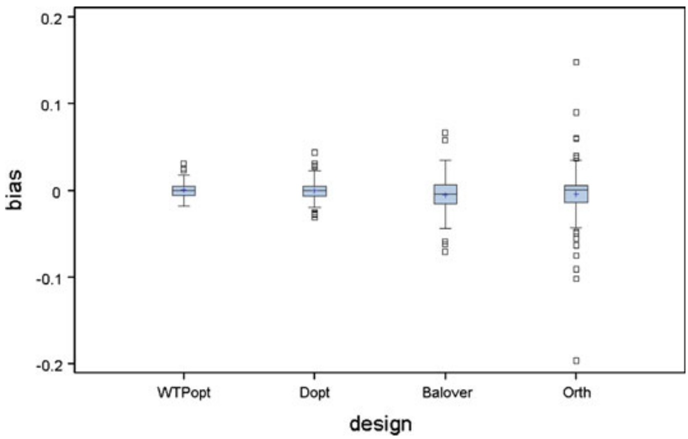

Fig. 2 Bias B_{WTP} of the WTP estimates for the different designs assuming a correct prior distribution

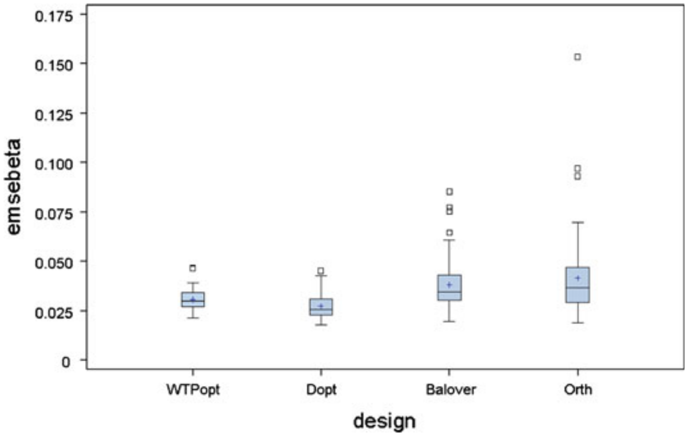

Fig. 3 EMSE_β values for the different designs assuming correct prior information

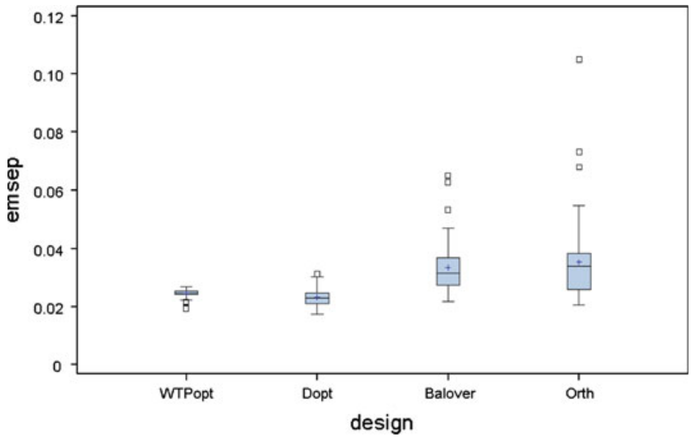

Fig. 4 EMSE_p values for the different designs assuming correct prior information

Figure 3 shows the 75 average EMSE_β values obtained from each of the 75 $\boldsymbol{\beta}$ vectors randomly drawn from the prior distribution. The box plots clearly indicate that the Bayesian D- and WTP-optimal designs produce substantially more precise estimates for the utility coefficients than the standard designs, but that the difference between the WTP-optimal and the D-optimal design is negligible. This means that focusing on precise WTP estimation when constructing a choice design does not come at a large cost in terms of the precision of the estimation of the utility coefficient vector $\boldsymbol{\beta}$.

Finally, Fig. 4 displays the prediction accuracy of the different designs as represented by the EMSE_p values. It can be seen that the D-optimal design leads to the most precise predictions, followed by the WTP-optimal design. The box plots clearly indicate that the two other benchmark designs result in considerably less precise predictions.

As a conclusion, we can say that the Bayesian WTP-optimal design leads to the most accurate marginal WTP estimates and to estimates for $\boldsymbol{\beta}$ that are nearly as precise as those obtained from the Bayesian D-optimal design if the prior information about the unknown parameters is reasonably correct. An additional advantage of the WTP-optimal design is that it leads to precise predictions as well. Our results also show that the nearly orthogonal design performs poorly compared to the Bayesian optimal designs.

4.3.3 Design Performance Under Incorrect Priors

In the previous section, we studied the relative performance of a Bayesian WTP-optimal design assuming that the prior distribution on $\boldsymbol{\beta}$ used to create the design contains reasonably correct information on the utility coefficients. In this section, however, we study the performance of the four competing designs in a scenario where the consumers' preferences are weaker or even counter to what was anticipated when constructing the design, and in a scenario where the consumers' preferences are stronger than expected. An earlier study by Ferrini and Scarpa (2007), among others, has indicated that the performance of Bayesian D-optimal designs depends on the correctness of the prior information used to construct the designs. It is, of course, necessary to re-address this issue for the Bayesian WTP-optimality criterion.

In a first scenario, we assume that the consumers' preferences are less pronounced than anticipated and can even be counter to what was expected when constructing the design. For that purpose, we randomly drew 75 $\boldsymbol{\beta}$ vectors from the 6-dimensional normal distribution with mean [0, 0, 0, 0, 0, −0.7] and variance-covariance matrix

$$\begin{pmatrix} \mathbf{I}_M & \mathbf{0}_{M\times 1} \\ \mathbf{0}_{1\times M} & 0.05 \end{pmatrix}.$$

We then used each randomly drawn $\boldsymbol{\beta}$ vector to simulate 1000 data sets for each of the four competing designs. For reasons of brevity of the paper, the following discussion only focuses on the EMSE_{WTP} and EMSE_{β} evaluation criteria.

Figure 5 shows the logarithms of the 75 average EMSE_{WTP} values of the four competing designs. The results show that, even if the consumers' preferences deviate from the prior distribution, the Bayesian WTP-optimal design measures the marginal WTPs more accurately than the other designs. The most striking result is again that the standard designs produce substantially more and larger extreme estimates of the marginal WTPs than the Bayesian WTP-optimal design and the D-optimal design. Figure 6 visualizes the EMSE_{β} values for

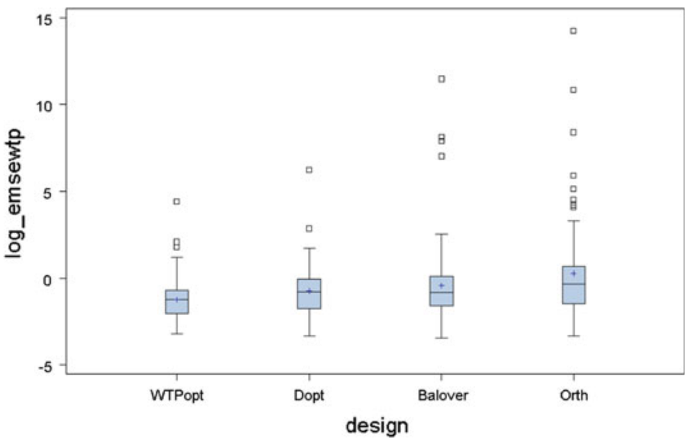

Fig. 5 Log (EMSE_{WTP}) values for the different designs assuming incorrect prior information

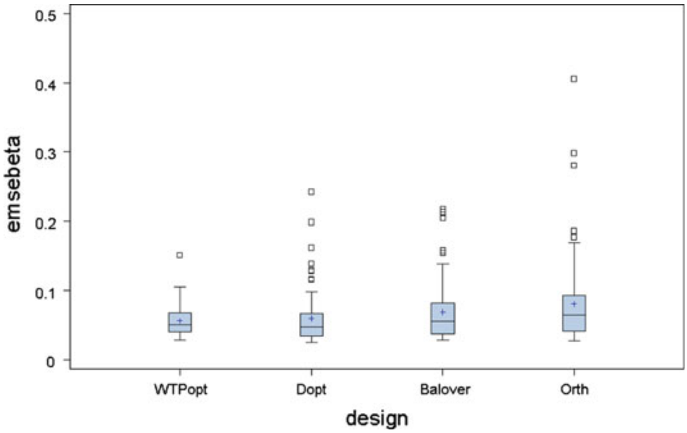

Fig. 6 EMSE_{β} values for the different designs assuming incorrect prior information

the different designs. The figure shows that the Bayesian WTP-optimal design yields more precise estimates of the utility coefficient vector $\boldsymbol{\beta}$ than the standard designs, and estimates that are nearly as precise as those from the Bayesian D-optimal design.

In a second scenario with incorrect prior information, we studied the relative performance of the WTP-optimal design when the consumers' preferences were more pronounced, or stronger, than anticipated. In this scenario, it was also assumed that the respondents were more sensitive to changes in price. This is reflected in the parameters of the distribution we drew $\boldsymbol{\beta}$ vectors from: a normal distribution with mean $[-1, 0, -1, 0, -1, -1]$ and variance-covariance matrix

$$\begin{pmatrix} 0.25\mathbf{I}_M & \mathbf{0}_{M\times 1} \\ \mathbf{0}_{1\times M} & 0.05 \end{pmatrix}.$$

It turned out that the results for this scenario were very similar to those in Figs. 5 and 6. Therefore, we do not show any detailed results for this scenario.

In summary, the results obtained from this simulation study clearly show that the Bayesian WTP-optimal design produces more accurate marginal WTP estimates than any of the other designs, including the Bayesian D-optimal one. This increased accuracy is to a large extent insensitive to the specification of the prior information used to construct the design. Moreover, and this is a novel result, the Bayesian WTP-optimal design yields considerably smaller and fewer extreme values for the marginal WTP estimates than the benchmark designs. This is an important contribution in solving the problem of unrealistically large marginal WTP estimates. The Bayesian WTP-optimal design also offers two additional advantages. First, it results in parameter estimates almost as precise as the Bayesian D-optimal design, suggesting that precision in estimation of the marginal WTPs does not come at a large loss in efficiency of the utility coefficient estimates. Second, the WTP-optimal design also has a good predictive performance.

5 The Willingness to Donate for Environmental Projects

In this section, we investigate the practical advantages of using WTP-optimal designs by revisiting an example described in Carlsson and Martinsson (2001). Based on the utility coefficients from the original study, simulated data were used to compare locally and Bayesian WTP-optimal designs with the original design of the Carlsson and Martinsson (2001) study in terms of the accuracy of the WTP estimates.

The example involves a choice experiment to value the willingness to donate for environmental projects. Three attributes were included in the study: the amount of money the respondents received, the donation they gave to an environmental project and the type of environmental project. In the choice experiment, every respondent had to make a trade-off between the money he/she received and the donation he/she made to support an environmental project. The amount of money the respondent received was 35, 50 or 65 Swedish Krona, whereas the possible donations amounted to 100, 150 or 200 Krona. The donations were intended to support a project in a rain forest, the Mediterranean Sea or the Baltic Sea. The authors used dummy coding for the type of project, and the Baltic Sea project was taken as the reference category.

The experiment involved 14 choice sets of two alternatives and used 35 respondents, yielding 490 observations in total. In Carlsson and Martinsson (2001) a locally D-optimal design was used, based on the information of a pilot study which suggested that the marginal willingness to donate for environmental projects was around five Krona. As the pilot study did not allow the estimation of the utility coefficients of the environmental projects, these were set to zero to generate the locally D-optimal design.

As an alternative to the locally D-optimal design, we propose locally and Bayesian WTP-optimal designs. The designs we discuss below were computed based on the utility coefficient vector [0.2, 1, 0, 0], which is in accordance with the information coming from the pilot study. The elements of the vector correspond to the utility coefficients of the money the respondents received, the donation and the environmental projects, respectively. For computing the Bayesian WTP-optimal design, we used a normal prior distribution with mean [0.2, 1, 0, 0] and variance-covariance matrix $0.5 \cdot \mathbf{I}_4$, where $\mathbf{I}_4$ is the four-dimensional identity matrix, to reflect the prior uncertainty about the utility coefficients. The alternating-sample algorithm, described in Kessels et al. (2009), was used to find the Bayesian WTP-optimal design.

Using the estimated model reported in Carlsson and Martinsson (2001) and summarized in Table 3, we simulated $R = 1,500$ data sets. For each of the 1,500 simulated data sets, an estimate of the marginal willingness to donate, the equivalent of the marginal WTP in this study, was computed. We did so for the locally D-optimal design used in Carlsson and Martinsson (2001), the locally WTP-optimal design and the Bayesian WTP-optimal design. To evaluate the three designs, we display the 1,500 marginal WTP estimates resulting from the use of the different designs graphically in Fig. 7. As in the simulation study above, we also computed the expected mean squared error EMSE_{WTP} and the bias B_{WTP}, and we counted the number of outliers in the marginal WTP estimates for each of the designs. Note that, in this case, the evaluation criteria are computed for only one marginal WTP value.

Table 3 Utility coefficient estimates, standard errors and WTP estimate for the original willingness-to-donate study

Variable	Coefficient	St. error
Money	0.033	0.010
Donation	0.021	0.003
Mediterranean	−0.885	0.148
Rainforest	−0.088	0.145
Marginal WTP donation	0.636	

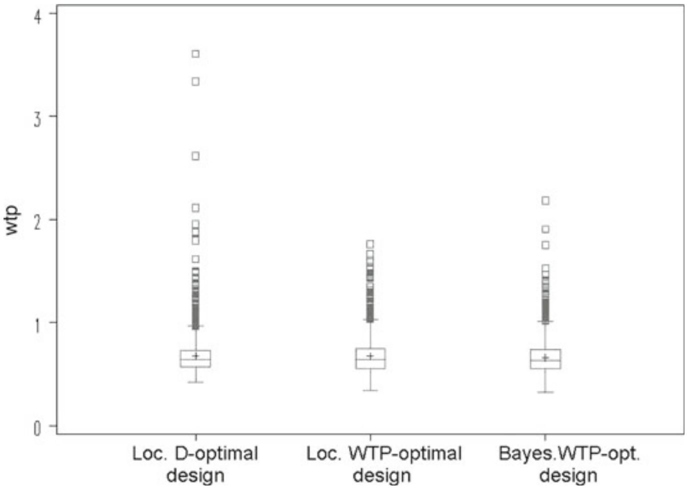

Fig. 7 Marginal WTP estimates from a locally *D*-optimal, a locally WTP-optimal and a Bayesian WTP-optimal design using the prior information of the pilot study for 1,500 simulated data sets

Table 4 Comparison of a locally D-optimal design and locally and Bayesian WTP-optimal designs using the information of a pilot study based on R =1,500 simulated data sets

Criterion	Locally D-optimal	Locally WTP-optimal	Bayesian WTP-optimal
EMSE_{WTP}	0.044	0.033	0.029
B_{WTP}	0.043	0.035	0.023
# Outliers	8	0	1

The box plots in Fig. 7 clearly show that the use of locally and Bayesian WTP-optimal designs results in fewer and smaller outlying estimates for the marginal WTP. This is clearly shown in Table 4, where the simulation results are summarized. Table 4 also shows that the EMSE_{WTP} values and the bias B_{WTP} for the WTP-optimal designs are substantially smaller than those for the locally D-optimal design.

The results for the prior utility coefficient estimate [0.2, 1, 0, 0] are representative for those obtained for other prior point estimates that take into account the results from the pilot study. We also found that the Bayesian WTP-optimal design approach still outperforms the other approaches in terms of the accuracy of the marginal WTP estimates if the prior information about the unknown parameters is to a substantial extent incorrect.

6 Discussion

In this paper, following Kanninen (1993) and Alberini (1995), we apply a c-optimality criterion to create optimal designs for conjoint choice experiments to estimate marginal WTP values accurately, and we refer to the resulting designs as WTP-optimal designs. We subject the Bayesian WTP-optimal designs to a series of comparisons with other more conventional designs. We use simulation and alternatingly assume correct and incorrect prior information about the utility coefficients generating the true responses. The results show that the Bayesian WTP-optimal designs consistently produce marginal WTP estimates that are substantially more accurate than those produced by other designs, including the Bayesian D-optimal designs, which under correct information were found to dominate more conventional designs in similar comparisons as reported in Ferrini and Scarpa (2007). Our results remain valid even if the prior information is not entirely correct. Importantly, the Bayesian WTP-optimal designs lead to smaller and fewer extreme values for the marginal WTP estimates. Finally, we note that the advantages offered by the Bayesian WTP-optimal design come at a negligible cost in terms of loss of efficiency in the utility coefficient estimates when compared to results obtained from a Bayesian D-optimal design. Finally, it was shown that the WTP-optimal design has good predictive performance. The WTP-optimality criterion therefore appears to be a valuable criterion in experimental design for conjoint choice experiments undertaken for the purpose of attribute valuation.

Design principles as orthogonality, D-optimality, level balance or utility balance do not seem to be appropriate design criteria to construct designs to obtain the most precise marginal WTP estimates. In summary, simulation results tell us that the Bayesian WTP-optimal designs clearly outperform these classical design principles in two ways when the goal of the choice experiment is to estimate the marginal WTPs. First, the WTP-optimal designs result in the most accurate WTP-estimates. Second, they significantly reduce the number and size of outlying WTP estimates compared to the other designs. Obtaining accurate WTP estimates requires the minimization of the variance of a non-linear function of utility coefficients in a non-linear model. Because of the non-linearity of the choice models and the function of interest, there is no theoretical justification for creating choice designs based on orthogonality and attribute level balance considerations. In fact, the most informative choice designs are not level balanced and not orthogonal. Moreover, it turns out that orthogonal designs and nearly orthogonal designs may result in serious estimation problems. In this article, this was visible from the unrealistically large WTP-estimates that we sometimes obtained.

In this paper, we constructed designs in so-called preference space: this means that the utility of an alternative is expressed in terms of the utility coefficients. However, the utility might also be expressed in terms of the WTPs and the price coefficients. This can easily be done by a reparameterization of the random utility model, and more specifically, by multiplying and dividing every term of the utility expression in preference space by the price coefficient (Train and Weeks 2005; Scarpa et al. 2008). The resulting utility is then expressed in WTP-space. The exercise to construct Bayesian designs in WTP-space minimizing the variance of the WTP estimates is a possible alternative to the WTP-optimal designs in preference space. This approach is subject of ongoing research, some preliminary results of which can be found in Vermeulen et al. (2009). As a future research question, we also mention the possibility to develop a design criterion for experiments which are focused on an accurate measurement of the compensating variation which is a complementary welfare measure to marginal WTP.

Acknowledgments The first author's research was funded by the project G.0611.05 of the Fund for Scientific Research Flanders.

Appendix

See Table 5.

Table 5 The four competing designs used in this study: the Bayesian WTP-optimal design, the Bayesian D-optimal design, the balanced overlap design and the nearly orthogonal design

Choice set	Alternative	WTP-opt.				D-opt.				Nearly orth.				Bal. overl.			
		1	2	3	p	1	2	3	p	1	2	3	p	1	2	3	p
1	1	2	3	1	1	1	3	1	1	3	2	1	2	3	3	2	1
	2	1	2	1	2	3	1	2	1	2	1	2	1	1	2	1	2
	3	2	3	1	2	2	2	2	2	1	3	1	1	2	1	2	2
2	1	1	3	1	1	2	1	1	1	2	2	1	1	1	3	1	1
	2	2	2	1	2	3	2	2	2	1	1	2	2	3	2	2	1
	3	1	3	1	2	1	3	2	1	3	3	2	2	2	1	1	2
3	1	1	2	2	2	3	3	2	2	3	1	1	1	1	2	2	2
	2	2	2	2	1	2	2	1	1	2	3	1	2	2	1	1	1
	3	2	1	2	2	1	1	2	1	1	2	2	2	3	3	1	2
4	1	3	1	2	1	2	1	2	1	3	3	2	1	3	1	1	1
	2	2	3	1	2	1	2	2	2	2	2	2	1	2	2	1	1
	3	3	2	2	2	2	2	1	2	1	1	1	2	2	3	2	2
5	1	1	3	2	2	3	1	1	1	3	1	1	1	3	2	1	1
	2	3	2	1	2	2	3	2	2	1	2	2	1	3	1	2	2
	3	2	2	2	1	1	2	1	1	2	3	1	2	1	3	2	2
6	1	1	1	2	1	3	3	1	2	2	1	2	2	1	2	1	2
	2	2	1	1	1	1	1	2	2	3	2	2	2	1	1	2	1
	3	2	2	2	2	2	2	2	1	1	3	1	1	2	3	2	1
7	1	2	3	2	2	2	1	1	2	2	2	1	2	1	3	2	1
	2	1	1	2	1	1	2	2	2	1	1	1	2	3	2	1	2
	3	1	3	1	1	3	3	1	1	3	3	2	1	2	1	2	1
8	1	3	3	1	1	2	2	2	1	2	1	2	1	1	1	1	1
	2	3	3	1	2	1	3	2	2	3	2	1	2	2	2	2	2
	3	1	1	2	2	3	1	1	2	1	3	2	1	2	3	1	2
9	1	3	1	1	2	2	1	2	2	2	3	2	2	2	3	1	2
	2	1	2	2	2	3	2	1	1	1	2	1	1	3	2	1	1
	3	2	1	2	1	1	3	1	2	3	1	1	1	3	1	2	1
10	1	1	3	2	1	3	1	2	2	2	2	1	1	3	2	1	2
	2	2	1	1	1	1	2	1	1	1	3	2	2	1	1	2	2
	3	3	3	2	2	2	3	1	1	3	1	2	2	1	3	2	1
11	1	2	1	2	2	2	3	1	2	3	3	1	2	2	2	2	1
	2	1	2	2	1	3	2	1	2	2	1	2	1	2	3	2	2
	3	1	3	1	2	1	1	2	1	1	2	2	1	1	1	1	1
12	1	3	1	2	2	2	1	2	2	1	1	1	2	3	2	2	1
	2	3	2	1	1	1	1	1	2	2	3	1	2	3	3	1	2
	3	2	2	1	2	1	2	2	1	3	2	2	1	1	1	1	2

The code to develop Bayesian WTP-optimal designs is available upon request

References

Adamowicz W, Boxall P, Williams M, Louviere J (1998) Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. Am J Agric Econ 80:64–75

Alberini A (1995) Optimal designs for discrete choice contingent valuation surveys: single-bound, double-bound, and bivariate models. J Environ Econ Manage 28:287–306

Atkinson AC, Donev AN (1992) Optimum experimental designs. Clarendon Press, Oxford

Atkinson AC, Haines LM (1996) Designs for nonlinear and generalized linear models. In: Ghosh S, Rao CR (eds) Handbook of statistics 13: design and analysis of experiments. Elsevier, Leiden, pp 437–475

Baiocchi G (2005) Monte carlo methods in environmental economics. In: Scarpa R, Alberini A (eds) Applications of simulation methods in environmental and resource economics, chapter 16. Springer, Dordrecht, pp 317–340

Banfi S, Farsi M, Filippini M, Jacob M (2008) Willingness to pay for energy-saving measures in residential buildings. Energy Econ 30:503–516

Bliemer MCJ, Rose JM (2010) Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Trans Res B 44:720–734

Boxall PC, Adamowicz WL (2002) Understanding heterogeneous preferences in random utility models: a latent class approach. Environ Resource Econ 23:421–446

Brau R, Cao D (2006) Uncovering the macrostructure of tourists' preferences. A choice experiment analysis of tourism demand of Sardinia. Note di Lavoro della Fondazione Eni Enrico Mattei

Carlsson F, Martinsson P (2001) Do hypothetical and actual marginal WTP differ in choice experiments?. J Environ Econ Manage 41:179–192

Carson R, Louviere J, Wasi N (2009) A cautionary note on designing discrete choice experiments: a comment on Lusk and Norwood's "Effect on experiment design on choice-based conjoint valuation estimates". Am J Agric Econ 91:1056–1063

Ferrini S, Scarpa R (2007) Designs with a-priori information for nonmarket valuation with choice experiments: a Monte Carlo study. J Environ Econ Manage 53:342–363

Hearne R, Salinas Z (2002) The use of choice experiments in the analysis of tourist preferences for ecotourism development in Costa Rica. J Environ Manage 65:153–163

Hensher D, Sullivan C (2003) Willingness to pay for road curviness and road type. Trans Res D 8:139–155

Hole AR (2007) A comparison of approaches to estimating confidence intervals for willingness to pay measures. Health Econ 16:827–840

Huber J, Zwerina K (1996) The importance of utility balance in efficient choice designs. J Market Res 33:307–317

Kanninen B (1995) Bias in discrete response contingent valuation. J Environ Econ Manage 28:114–125

Kanninen BJ (1993) Optimal experimental design for double-bounded dichotomous choice contingent valuation. Land Econ 69:138–146

Kessels R, Goos P, Vandebroek M (2006) A comparison of criteria to design efficient choice experiments. J Market Res 43:409–419

Kessels R, Jones B, Goos P, Vandebroek M (2008) Recommendations on the use of Bayesian optimal designs for choice experiments. Qual Reliab Eng Int 24:737–744

Kessels R, Jones B, Goos P, Vandebroek M (2009) An efficient algorithm for constructing Bayesian optimal choice designs. J Bus Econ Stat 27:279–291

Kimenju S, Morawetz U, Groote HD (2005) Comparing contingent valuation methods, choice experiments and experimental auctions in soliciting consumer preference for maize in Western Kenya: preliminary results, Paper prepared for presentation at the African Econometric Society 10th annual conference on econometric modeling in Africa, Nairobi, Kenya

Kuhfeld W, Tobias R, Garratt M (1994) Efficient experimental designs with marketing applications. J Market Res 31:545–557

Louviere J, Eagle T (2006) Confound it! That pesky little scale constant messes up our convenient assumptions! in Sawtooth Software Conference Proceedings; Sequem, Washington: Sawtooth Software, pp 211–228

Louviere JJ, Street D, Carson R, Ainslie A, DeShazo J, Cameron T, Hensher D, Kohn R, Marley A (2002) Dissecting the random component of utility. Market Lett 13:177–193

Lusk JL, Norwood FB (2005) Effect of experimental design on choice-based conjont valuation estimates. Am J Agric Econ 87:771–785

Lusk JL, Norwood FB (2009) A cautionary note on the design of discrete choice experiments: reply. Am J Agric Econ 91:1064–1066

Lusk JL, Roosen J, Fox JA (2003) Demand for beef from cattle administered growth hormones or fed genetically modified corn: a comparison of consumers in France, Germany, the U.K. and the U.S. Am J Agric Econ 85:16–29

Morey E, Rossmann K (2003) Using stated-preference questions to investigate variations in willingness to pay for preserving marble monuments: classic heterogeneity, random parameters, and mixture models. J Cultural Econ 27:215–229
Mtimet N, Albisu L (2006) Spanish wine consumer behavior: a choice experiment approach. Agribusiness 22:343–362
Nyquist H (1992) Optimal designs of discrete response experiments in contingent valuation studies. Rev Econ Stat 74:559–563
Ruto E, Garrod G, Scarpa R (2008) Valuing animal genetic resources: a choice modeling application to indigenous cattle in Kenya. Agric Econ 38:89–98
Ryan M (2004) A comparison of stated preference methods for estimating monetary values. Health Econ 13:291–296
Sammer K, Wüstenhagen R (2006) The influence of eco-labelling on consumer behaviour—results of a discrete choice analysis for washing machines. Bus Strategy Environ 15:185–199
Sándor Z, Wedel M (2001) Designing conjoint choice experiments using managers' prior beliefs. J Market Res 38:430–444
Scarpa R, Campbell D, Hutchinson WG (2007) Benefit estimates for landscape improvements: sequential Bayesian design and respondents' rationality in a choice experiment. Land Econ 83:617–634
Scarpa R, Rose J (2008) Design efficiency for non-market valuation with choice modelling: how to measure it what to report and why. Aust J Agric Resource Econ 52:253–282
Scarpa R, Thiene M, Train K (2008) Utility in WTP space: a tool to address confounding random scale effects in destination choice to the Alps. Am J Agric Econ 90:994–1010
Sonnier G, Ainslie A, Otter T (2007) Heterogeneity distributions of willingness-to-pay in choice models. Quant Market Econ 5(3):313–331
Swait J, Adamowicz W (2001) The influence of task complexity on consumer choice: a latent class model of decision strategy switching. J Consumer Res 28:135–148
Swait J, Louviere J (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models. J Market Res 30:305–314
Train K (2003) Discrete choice methods with simulation. Cambridge University Press, New York
Train K, Weeks M (2005) Discrete choice models in preference space and willingness-to-pay space. In: Scarpa R, Alberini A (eds) Applications of simulation methods in environmental and resource economics. Springer, Dordrecht
Vermeulen B, Goos P, Scarpa R, Vandebroek M (2009) Design criteria to develop choice experiments to measure the WTP accurately. Research Report KBI_0816, Department of Decision Sciences and Information Management, Katholieke Universiteit Leuven, 29 pp
Yu J, Goos P, Vandebroek M (2009a) Efficient conjoint choice designs in the presence of respondent heterogeneity. Market Sci 28:122–135
Yu J, Goos P, Vandebroek M (2009b) Individually adapted sequential Bayesian conjoint-choice designs in the presence of consumer heterogeneity. Research Report KBI_0902, Department of Decision Sciences and Information Management, Katholieke Universiteit Leuven, 15 pp