



Design criteria to develop choice experiments to measure the WTP accurately

B. Vermeulen, P. Goos, R. Scarpa and M. Vandebroek

DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

Design criteria to develop choice experiments to measure the WTP accurately

Bart Vermeulen¹

Faculty of Business and Economics, Katholieke Universiteit Leuven
Naamsestraat 69, B-3000 Leuven, Belgium
Email: bart.vermeulen@econ.kuleuven.be
Tel: +32 (0)16 32 69 63
Fax: +32 (0)16 32 66 24

Peter Goos

Faculty of Applied Economics, Universiteit Antwerpen
Prinsstraat 13, B-2000 Antwerpen, Belgium
Email: peter.goos@ua.ac.be
Tel: +32 (0)3 220 40 59
Fax: +32 (0)3 220 48 17

Riccardo Scarpa

Economics Department, University of Waikato
Private Bag 3105, Hamilton 3240, New Zealand
Email: rscarpa@waikato.ac.nz
Tel: +64 (0) 7 838 4045
Fax: +64 (0) 7 838 4331

Martina Vandebroek

Faculty of Business and Economics & Leuven Statistics Research Centre, Katholieke Universiteit Leuven
Naamsestraat 69, B-3000 Leuven, Belgium
Email: martina.vandebroek@econ.kuleuven.be
Tel: +32 (0)16 32 69 75
Fax: +32 (0)16 32 66 24

¹Corresponding author

Abstract

To measure the willingness-to-pay (WTP) accurately, Vermeulen et al. [2008] apply the c -optimality criterion to generate designs for conjoint choice experiments. This criterion is based on minimizing the sum of the variances of the WTP estimators approximated by the delta method. Designs generated based on this criterion lead to more accurate WTP estimates than the ones obtained by standard designs and reduce considerably the occurrence of extreme WTP estimates, although they do not exclude them. In this paper, other optimality criteria are considered to tackle this problem. We distinguish between criteria in preference space on the one hand and criteria in WTP -space on the other hand. In a simulation study and a numerical example, we compare the accuracy of the WTP and the utility coefficient estimates yielded by the designs based on these new criteria.

The absence of a market for commodities such as environmental goods (e.g. air and water quality) and health services (e.g. risk reduction programs in the context of disease prevention, treatments,...) makes their valuation difficult. As government interventions often deal with the allocation of these goods, the determination of their monetary value is important to decide on the appropriate level of supply. To assess the economic desirability of these policies and to align the commodities more closely with the preferences of the public, analysts must estimate the value of these nonmarket commodities and their characteristics. In these cost-benefit analyses, the concept of willingness-to-pay (*WTP*) has become an essential component.

Until recently, the contingent valuation method, which asks the respondents whether they are prepared to pay a certain amount for a hypothetical change in a nonmarket good, was the most frequently used method to elicit individuals' *WTP*. However, stated choice modelling, which originally found its use in market research, has become increasingly popular in nonmarket valuation. This technique involves presenting individuals with a number of scenarios or profiles each representing a commodity described in terms of its underlying characteristics or attributes. For each choice set, the respondents are asked to evaluate the presented alternatives and to choose their preferred one. Through the use of discrete choice models, these choice data inform the researcher which attributes are important for the individuals. Compared to the contingent valuation method, stated choice modelling allows the estimation of the value of single characteristics of a commodity and avoids 'yes-saying' of the respondents during the survey (Carlsson and Martinsson [2001]).

As the number of combinations of the attribute levels, and so the number of possible profiles, might be huge, guidance on profile selection and their bundling in choice sets is vital to obtain a maximum amount of information from the experiment. In this respect, a central role in stated choice studies is reserved to experimental design theory which uses various criteria to evaluate the quality of the assignment of attributes and their levels on the basis of the objectives set by the analyst. The selection of the correct criterion conditional on these objectives will lead to the most appropriate experimental design. As stated choice modelling found its origin in market research, the recent change of focus to nonmarket valuation implies some important challenges in the use of experimental design theory. This paper contributes to tackling one of those challenges, namely the

accurate estimation of the *WTP*.

The organization of the remainder of this paper is as follows. Section 1 briefly reviews the conditional logit model and the *WTP* concept. Section 2 reviews the existing criteria to develop experiments to measure the *WTP*. Section 3 addresses the different efficiency criteria. Their performance is compared in a simulation study in Section 4. Finally, Section 5 provides an application in health economics in which the designs generated based on the proposed criteria are evaluated.

1 The conditional logit model and the willingness-to-pay

Consider the situation where a respondent n has to choose the most preferred alternative in choice set k of size J . The utility of alternative j experienced by this decision maker can be written as

$$U_{nj k} = \beta_1 x_{1jk} + \dots + \beta_M x_{Mjk} + \varepsilon_{nj k} \quad (1)$$

or in vector notation $U_{nj k} = \mathbf{x}'_{jk} \boldsymbol{\beta} + \varepsilon_{nj k}$. As can be seen, the utility is assumed to be composed of two parts: a deterministic and a random component. The deterministic component consists of the vector $\boldsymbol{\beta}$ of utility coefficients expressing the importance of the attributes of alternative j in determining its utility and the vector \mathbf{x}_{jk} containing the levels of these attributes. The utility coefficients are assumed to be equal for all respondents. The random components $\varepsilon_{nj k}$ capture the unobserved influences on the respondents' utility and are assumed to be i.i.d. Gumbel distributed. The probability that respondent n chooses alternative j can therefore be written in the following form

$$p_{nj k} = \frac{\exp(\mathbf{x}'_{jk} \boldsymbol{\beta})}{\sum_{i=1}^J \exp(\mathbf{x}'_{ik} \boldsymbol{\beta})} \quad (2)$$

(Train [2003]).

The *WTP* expresses the willingness of individuals to obtain a change in an attribute of a commodity in monetary terms. From this perspective, the *WTP* corresponds to the marginal rate of substitution between an attribute m and the price p . In this way, the *WTP* measures the change in

the price p to compensate for a change in attribute m while all other attributes are held constant. In order to compute the WTP , one of the M attributes in \mathbf{x}_{jk} in utility expression (1) has to be the price. Mathematically, this trade-off between attribute m and the price p is given by

$$dU = \beta_m dx_m + \beta_p dp = 0, \quad (3)$$

which leads to the following definition of the WTP :

$$WTP_m \equiv \frac{dp}{dx_m} = -\frac{\beta_m}{\beta_p}. \quad (4)$$

The use of conjoint choice experiments to estimate the WTP has conquered an established role in valuation issues. Proof of this can be found in numerous studies which use this technique for that purpose. The applications are not any longer limited to marketing, in which the technique has been most frequently applied (e.g. Sammer and Wüstenhagen [2006]), but can also be retrieved in many other fields such as transport economics (e.g. Hensher and Sullivan [2003]), health economics (e.g. Ryan [2004] and Hole [2008]), environmental economics (e.g. Banfi et al. [2008]) and energy economics (e.g. Boxall and Adamowicz [2002]).

2 Measuring the WTP through conjoint choice experiments

The first approach to estimate the WTP accurately is described in Kanninen [1993] who focused on optimal designs for double-bounded dichotomous contingent valuation experiments. In single-bounded dichotomous contingent valuation experiments, the respondents are asked whether they are prepared to pay a certain amount for a hypothetical change in one or more product attributes. In order to analyze the data coming from this type of experiments, the conditional logit model considers the probability of answering affirmatively to the offered change in product attributes. In double-bounded experiments, the initial bid is followed by a higher bid if the answer to the first one was affirmative and lower otherwise. In this case, the probability of the occurrence of the sequence of responses is maximized in the estimation procedure. Kanninen [1993] compares the D -optimal design strategy, the c -optimal design strategy and a design strategy based on the so-called Fieller method to construct confidence intervals for a ratio of parameters. While D -optimal designs mini-

mize the confidence region of the estimated utility coefficients in the double-bounded logit model, c -optimal designs minimize the variance of a function of the estimated utility coefficients. The function of interest here was the sum of the variances of the WTP estimates approximated by the delta method. The third design was based on minimizing the length of the confidence interval of the WTP constructed by the Fieller method. As these criteria depend on the unknown utility coefficients that have to be estimated, their values are important for the design efficiency itself. This problem was circumvented by assuming a zero point estimate to determine the optimal design. The three design strategies were examined in terms of the variance of the WTP estimates and it turned out that c -optimal designs and the designs based on the Fieller method performed better than D -optimal designs. However, the difference between the three design strategies was small for the double-bounded logit model.

Also Alberini [1995] looked at the performance of several design strategies to precisely estimate the WTP by means of a contingent valuation survey. She considered the single-bounded, the double-bounded and the bivariate case. In contrast with the double-bounded case, the follow-up bid in the bivariate one does not have to be higher (lower) if the respondents answer affirmatively (negatively) to the first bid. When performing a bivariate contingent valuation survey, it is important that the follow-up bid does not come immediately after the initial bid to avoid confusion of the respondents because of a counter-intuitive follow-up bid. This bivariate concept reflects the idea that the respondents may revise their WTP during the survey. Assuming zero prior utility coefficients, Alberini considered three types of locally optimal designs: D -optimal designs, designs based on the Fieller method and variance-minimizing designs which are also known as A -optimal designs and minimize the variance of the WTP estimates. The results revealed a small difference in WTP estimation accuracy between these three designs, although not sufficiently large to rule out one of them.

A first approach to estimate the WTP by means of conjoint choice experiments is described in Ferrini and Scarpa [2007], who compared the precision of the WTP estimates yielded by shifted designs, locally D -optimal designs and Bayesian D -optimal designs. In shifted designs, the profiles of a choice set are generated by repeatedly increasing all attribute levels of a profile of a

K -point fractional factorial orthogonal starting design with one unit (if an attribute already is at its highest level, this level is changed to the lowest level admissible for that attribute). The number of times the attribute levels are increased in this fashion equals $J - 1$. Every profile of the starting design is used in this way to create a choice design with K choice sets. D -optimal designs minimize the generalized variance of the utility coefficients of the conditional logit model by minimizing the determinant of the variance-covariance matrix of their estimates. Again, the efficiency of a design depends on the utility coefficients which have to be estimated. Locally D -optimal designs circumvent this problem by using a point estimate for the utility coefficients as prior knowledge to determine the optimal design. Bayesian D -optimal designs assume a prior distribution on the utility coefficients to formally account for the uncertainty about their values. The results in Ferrini and Scarpa [2007] suggested that important improvements in the accuracy of WTP estimates can be achieved by using Bayesian D -optimal designs developed assuming an informative prior distribution. These improvements are positively correlated with the coefficients' magnitude. However, when assuming an uninformative prior, the most accurate WTP estimates are obtained using shifted designs.

In Rose and Scarpa [2008], a number of design criteria to generate designs for choice experiments, among others the D - and c -optimality criterion, were compared in terms of their relative efficiency. Like in Kanninen [1993], the c -optimality criterion consists of the sum of the variances of the WTP approximated by the delta method. This comparison revealed that c -optimal designs perform surprisingly well with respect to the different efficiency criteria. Remarkably, it appeared that D -optimal designs are also a valuable design option in terms of the c - and other optimality criterion considered in their study.

Finally, Vermeulen et al. [2008] provided a comparison between different design strategies in terms of the precision of the WTP and utility coefficient estimates. By means of a simulation study, three standard designs were compared with Bayesian D -optimal and Bayesian c -optimal designs. The three standard designs discussed in the study were random designs, orthogonal designs and designs exhibiting a limited attribute level overlap within a choice set. The study unveiled that Bayesian c -optimal designs outperform the other designs in terms of WTP estimation accuracy

if an informative prior is used. Bayesian c -optimal designs also yield the most accurate WTP estimates if the prior information is incorrect. Moreover, the results revealed that this type of design is an important step in reducing the frequently encountered problem of unrealistically large WTP estimates caused by a poorly estimated price coefficient. Finally, Bayesian c -optimal designs show the desirable characteristic of allowing the utility coefficients to be estimated almost as precisely as Bayesian D -optimal designs. Although this contribution answers several questions regarding Bayesian c -optimal designs, it does not completely rule out the occurrence of extreme WTP estimates. That is why further research is needed to develop a design criterion which results in accurately estimated WTP values and which avoids the occurrence of extreme estimates as much as possible.

3 Design criteria to estimate the WTP accurately

This section describes the different design criteria we consider to generate conjoint choice designs for estimating the WTP precisely. We distinguish between two categories of criteria: (i) criteria in preference space, which specifies the utility in terms of the utility coefficients of the attributes, and (ii) criteria in WTP -space which defines the utility of a commodity in terms of the WTP s and the price coefficient. The criteria in WTP -space are based on a reparameterization of the random utility model, which is obtained by multiplying and dividing every term of the utility expression in preference space by the price coefficient (Train and Weeks [2005] and Scarpa et al. [Forthcoming]). Henceforth, we indicate a design criterion in preference space by the superscript 'pref' and a design criterion in WTP -space by the superscript 'WTP'.

3.1 Design criteria in preference space

To estimate the WTP accurately, Rose and Scarpa [2008] and Vermeulen et al. [2008] applied the c^{pref} -optimality criterion which minimizes the sum of the variances of the WTP estimates approximated by the delta method. The approximate variance of the WTP estimate for attribute m can be obtained as follows:

$$\begin{aligned}\widehat{var}(\widehat{WTP}_m) &= \widehat{var}\left(-\frac{\hat{\beta}_m}{\hat{\beta}_p}\right) \\ &\approx \frac{1}{\hat{\beta}_p^2} \left(var(\hat{\beta}_m) - 2 \left(\frac{\hat{\beta}_m}{\hat{\beta}_p}\right) cov(\hat{\beta}_m, \hat{\beta}_p) + \left(\frac{\hat{\beta}_m}{\hat{\beta}_p}\right)^2 var(\hat{\beta}_p) \right).\end{aligned}\quad (5)$$

The efficiency of a design in terms of the c^{pref} -optimality criterion is then given by

$$c^{pref}\text{-error} = \sum_{m=1}^{M-1} \widehat{var}(\widehat{WTP}_m) = \widehat{var}(\widehat{\mathbf{W}}) \quad (6)$$

with $\widehat{\mathbf{W}}$ is a $(M-1)$ -dimensional vector consisting of the WTP estimates. The design minimizing the Bayesian version of the c^{pref} -error

$$c_b^{pref}\text{-error} = E_{\beta} \left[\sum_{m=1}^{M-1} \widehat{var}(\widehat{WTP}_m) \right] = \int_{\mathfrak{R}^M} \widehat{var}(\widehat{\mathbf{W}}) \pi(\beta) d\beta, \quad (7)$$

where $\pi(\beta)$ denotes the prior distribution of β , is referred to as the Bayesian c^{pref} -optimal design.

An undesirable feature of this design criterion is that its application may lead to the unsatisfactory outcome that a design is created by reducing the variances unevenly across the WTP estimates considered in the study: the WTP for one attribute can be estimated precisely while the WTP for an other attribute is not. Two criteria which might offer a solution to this drawback involve minimizing the maximum value of the variances of the WTP estimates and minimizing the variance of the variances of the WTP estimates. We refer to these criteria as the $minimax^{pref}$ optimality criterion and the $varmin^{pref}$ optimality criterion, respectively. The variance of the WTP in these two criteria is again approximated by the delta method.

The performance of a design in terms of the $minimax^{pref}$ optimality criterion is given by

$$minimax^{pref}\text{-error} = \max\{\widehat{var}(\widehat{WTP}_1), \widehat{var}(\widehat{WTP}_2), \dots, \widehat{var}(\widehat{WTP}_{M-1})\}. \quad (8)$$

To find the Bayesian $minimax^{pref}$ optimal design, the $minimax^{pref}$ -error has to be minimized

over the prior distribution. The Bayesian version of the $minimax^{pref}$ -error can be found in Table 1. Although this criterion implies that it will estimate the WTP with the largest variance in the most accurate way, it disregards the variance of the other WTP s possibly yielding relatively less accurate estimates.

An alternative criterion to overcome the shortcomings of the two previous ones might be the $varmin^{pref}$ optimality criterion, which minimizes the variance of the variances of the WTP estimates. This might avoid large values of the variance of the WTP estimates and so, prevents extreme values of the WTP itself. The corresponding error measuring the efficiency of a design in terms of this criterion is given by

$$varmin^{pref}\text{-error} = var\{\widehat{var}(\widehat{WTP}_1), \widehat{var}(\widehat{WTP}_2), \dots, \widehat{var}(\widehat{WTP}_{M-1})\}. \quad (9)$$

The performance of a design in terms of the Bayesian version of this criterion is then measured by $varmin_b^{pref}$ -error which is given in Table 1.

A fourth design criterion considered in preference space is based on the Fieller method to construct confidence intervals for a ratio of parameters. The confidence interval for the WTP constructed by this method is the set of WTP values which could have given rise to the observed choices of individuals with a specified probability $(1 - \alpha)$ (Wang [2000]). The Fieller confidence interval or fiducial interval for the WTP_m is then given by

$$\frac{\hat{\beta}_m - z_{\alpha/2}^2 covar(\hat{\beta}_m, \hat{\beta}_p)/\hat{\beta}_m \pm \Phi_m^{1/2}}{-\hat{\beta}_m + z_{\alpha/2}^2 var(\hat{\beta}_m)/\hat{\beta}_m}, \quad (10)$$

with

$$\Phi_m = (\hat{\beta}_m - z_{\alpha/2}^2 covar(\hat{\beta}_m, \hat{\beta}_p)/\hat{\beta}_p)^2 - (1/\hat{\beta}_p)^2 (\hat{\beta}_m^2 - z_{\alpha/2}^2 var(\hat{\beta}_m)) (\hat{\beta}_p^2 - z_{\alpha/2}^2 var(\hat{\beta}_p)) \quad (11)$$

and z being the value of the standard normal variate that cumulates 95% probability mass. The design criterion derived from the Fieller method minimizes the sum of the absolute values of Φ_m associated with the attributes under investigation, and so of the width of the intervals for the related WTP estimates. We consider the absolute value of Φ_m because it is possibly negative if

the variance for the estimated coefficients β_m and β_p is larger than 0.3 (Hole [2007]). If there are M attributes including the price, the error in terms of this criterion is defined as the sum of the width for the intervals for $M - 1$ WTP estimates

$$FID^{pref}\text{-error} = \sum_{m=1}^{M-1} |\Phi_m|. \quad (12)$$

The performance of a design in terms of the Bayesian *fiducial^{pref}* optimality criterion is then given by the FID_b^{pref} -error shown in Table 1.

A last design criterion in preference space follows the approach of Toubia and Hauser [2007]. They argued that managerial decisions are not always taken based on the utility coefficients themselves, but often on quantities which are functions of the utility coefficients. Standard design criteria, such as the D^{pref} -optimality criterion, do not necessarily yield accurate estimates of these quantities. That is why Toubia and Hauser [2007] developed the so-called M^{pref} -optimality criterion to accommodate this phenomenon. Following their approach, the asymptotic variance-covariance matrix of the WTP estimates is approximated by

$$\Omega_{WTP}(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{M}(\mathbf{X}'(\mathbf{P} - \mathbf{p}\mathbf{p}')\mathbf{X})^{-1}\mathbf{M}' \quad (13)$$

where the ij th element of \mathbf{M} is $\frac{\partial WTP_i}{\partial \beta_j}$. Based on this approximate variance-covariance matrix, the volume of the confidence region of the WTP estimates can be minimized. The error corresponding to this approach is defined as

$$M^{pref}\text{-error} = \det(\mathbf{M}(\mathbf{X}'(\mathbf{P} - \mathbf{p}\mathbf{p}')\mathbf{X})^{-1}\mathbf{M}') = \det(\Omega_{WTP}(\mathbf{X}, \boldsymbol{\beta})). \quad (14)$$

To develop Bayesian M^{pref} -optimal designs, the M_b^{pref} -error given in Table 1 has to be minimized.

To generate Bayesian optimal designs, the Bayesian version of the expressions (6), (8), (9), (12) and (14), which can be found in Table 1, has to be minimized. It has to be pointed out that no analytical expressions can be found for these Bayesian criteria. Consequently, the integral in these expressions has to be approximated using a number of draws from the prior distribution and averaging the corresponding error over these draws. The coordinate exchange algorithm was the

procedure applied to search for the optimal design (Meyer and Nachtsheim [1995]).

3.2 Design criteria in WTP -space

Multiplying and dividing every term in utility expression (1) by the price coefficient β_p results in an expression in which the utility is defined in terms of the WTP and the price coefficient

$$\begin{aligned} U_{njk} &= \frac{\beta_1}{\beta_p} \beta_p x_{1jk} + \dots + \frac{\beta_{M-1}}{\beta_p} \beta_p x_{(M-1)jk} + \beta_p \frac{\beta_p}{\beta_p} p_{jk} + \varepsilon_{njk} \\ &= \beta_p WTP_1 x_{1jk} + \dots + \beta_p WTP_{M-1} x_{(M-1)jk} + \beta_p p_{jk} + \varepsilon_{njk}. \end{aligned} \quad (15)$$

Analyzing choice data by a conditional logit model defined in WTP -space avoids unrealistic WTP estimates but comes at a cost of a reduced fit (Train and Weeks [2005] and Scarpa et al. [Forthcoming]). The information matrix \mathbf{I}_{WTP} on the WTP and the price coefficient estimates for a choice set k can then be written as

$$\mathbf{I}_{WTP} = \begin{bmatrix} \boldsymbol{\kappa} & \boldsymbol{\phi} \\ \boldsymbol{\phi}' & \tau \end{bmatrix} \quad (16)$$

with

$$\boldsymbol{\kappa} = \beta_p \mathbf{X}'_{min} (\mathbf{P} - \mathbf{p}\mathbf{p}') \mathbf{X}_{min} \beta_p, \quad (17)$$

$$\boldsymbol{\phi} = \beta_p \mathbf{X}'_{min} (\mathbf{P} - \mathbf{p}\mathbf{p}') \mathbf{X} \cdot \mathbf{WTP}, \quad (18)$$

and

$$\tau = \mathbf{WTP}' \cdot \mathbf{X}' (\mathbf{P} - \mathbf{p}\mathbf{p}') \mathbf{X} \cdot \mathbf{WTP} \quad (19)$$

where the $(J \times (M - 1))$ -dimensional matrix \mathbf{X}_{min} denotes the design matrix containing the attributes of the alternatives in choice set k except for the price, \mathbf{X} is the $(J \times M)$ -dimensional design matrix including the price attribute and the M -dimensional vector \mathbf{WTP} contains the utility coefficients divided by the price coefficient. Consequently, the matrix $\boldsymbol{\kappa}$ stands for the variance and the covariance of the WTP estimates, $\boldsymbol{\phi}$ for the covariance of the WTP and price coefficient estimates and finally, τ represents the variance of the estimated price coefficient.

Given that the Fisher information matrix \mathbf{I}_{WTP} is inversely proportional to the variance-covariance matrix \mathbf{V}_{WTP} on the WTP and the price coefficient estimates, D^{WTP} -optimal designs minimize the generalized variance of the WTP and price coefficient estimates measured by the determinant of \mathbf{V}_{WTP} . The performance of a design in terms of the D^{WTP} -optimality criterion is expressed by

$$D^{WTP}\text{-error} = (\det(\mathbf{V}_{WTP}(\mathbf{X}, \boldsymbol{\beta})))^{\frac{1}{M}}. \quad (20)$$

Minimizing the Bayesian version of the D^{WTP} -error, which is given in Table 1, leads to the Bayesian D^{WTP} -optimal design.

Besides minimizing the determinant of \mathbf{V}_{WTP} , one can also minimize its trace which leads to A^{WTP} -optimal designs. The trace is then used as a measure for the performance of a design and is given by

$$A^{WTP}\text{-error} = \text{tr}(\mathbf{V}_{WTP}(\mathbf{X}, \boldsymbol{\beta})). \quad (21)$$

The Bayesian version of the A^{WTP} -error given in Table 1 allows us to obtain the Bayesian A^{WTP} -optimal design.

The expression for the Fisher information matrix \mathbf{I}_{WTP} allows to formulate a criterion minimizing the variance of the variances of the WTP estimates, henceforth called the $varmin^{WTP}$ optimality criterion. The performance of a design is then measured by

$$varmin^{WTP}\text{-error} = \text{var}\{\text{var}(\widehat{WTP}_1), \text{var}(\widehat{WTP}_2), \dots, \text{var}(\widehat{WTP}_{M-1})\}. \quad (22)$$

Minimizing this error over the prior distribution gives the Bayesian $varmin^{WTP}$ optimal design. The difference with the $varmin^{pref}$ optimality criterion in preference space is in the way the variance of the WTP estimate is calculated. The elements on the main diagonal of the variance-covariance matrix \mathbf{V}_{WTP} are the variances of the WTP estimates, and so, need not be approximated by the delta method. Minimizing the Bayesian version of the expressions (20), (21) and (22) leads to Bayesian optimal designs in WTP -space. For that purpose, we use the same procedure as in preference space.

Bayesian optimality criteria to estimate the WTP accurately	
Criteria in preference space	
$c_b^{pref}:$	$c_b^{pref} = \int_{\mathfrak{R}^M} \widetilde{var}(\widehat{\mathbf{W}}) \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}$
$minimax^{pref}:$	$minimax_b^{pref} = \int_{\mathfrak{R}^M} \max\{\widetilde{var}(\widehat{WTP}_1), \dots, \widetilde{var}(\widehat{WTP}_{M-1})\} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}$
$varmin^{pref}:$	$varmin_b^{pref} = \int_{\mathfrak{R}^M} \text{var}\{\widetilde{var}(\widehat{WTP}_1), \dots, \widetilde{var}(\widehat{WTP}_{M-1})\} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}$
$fiducial^{pref}:$	$FID_b^{pref} = \int_{\mathfrak{R}^M} \sum_{m=1}^{M-1} \Phi_m \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}$
$M^{pref}:$	$M_b^{pref} = \int_{\mathfrak{R}^M} \det(\boldsymbol{\Omega}_{WTP}(\mathbf{X}, \boldsymbol{\beta})) \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}$
Criteria in WTP- space	
$D^{WTP}:$	$D_b^{WTP} = \int_{\mathfrak{R}^M} (\det(\mathbf{V}_{WTP}(\mathbf{X}, \boldsymbol{\beta})))^{\frac{1}{M}} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}$
$A^{WTP}:$	$A_b^{WTP} = \int_{\mathfrak{R}^M} \text{tr}(\mathbf{V}_{WTP}(\mathbf{X}, \boldsymbol{\beta})) \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}$
$varmin^{WTP}:$	$varmin_b^{WTP} = \int_{\mathfrak{R}^M} \text{var}\{\text{var}(\widehat{WTP}_1), \dots, \text{var}(\widehat{WTP}_{M-1})\} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}$
Benchmark criteria	
D^{pref}	$D_b^{pref} = \int_{\mathfrak{R}^M} (\det(\mathbf{V}(\mathbf{X}, \boldsymbol{\beta})))^{\frac{1}{M}} \pi(\boldsymbol{\beta}) d(\boldsymbol{\beta})$
A^{pref}	$A_b^{pref} = \int_{\mathfrak{R}^M} \text{trace}(\mathbf{V}(\mathbf{X}, \boldsymbol{\beta})) \pi(\boldsymbol{\beta}) d(\boldsymbol{\beta})$

Table 1: Overview of the Bayesian design criteria

In the next section, these designs will be evaluated in terms of their ability to estimate the WTP and the utility coefficients accurately. Bayesian D -optimal and A -optimal designs in preference space, henceforth denoted as Bayesian D^{pref} -optimal and A^{pref} -optimal designs, are considered as benchmark designs. We refer the reader to Sándor and Wedel [2001], Sándor and Wedel [2005] and Kessels et al. [2006a] for further details on how to create these designs.

4 Evaluation of the designs

In this section, we first give more details concerning the development of the designs. Then, after explaining two evaluation criteria measuring the accuracy of the WTP estimates and the estimates

of the utility coefficients, we investigate the performance of the designs in terms of these measures by means of a simulation study assuming correct and incorrect prior information.

4.1 Developing the designs

The experiment considered consists of eight choice sets, each containing three alternatives. Each alternative is described by means of three attributes which are all effects-type coded: two of the attributes have three levels and the third one only takes two possible levels. Besides these three attributes, the price is also included as an attribute. The price has three levels in the study: these are linearly coded as 1, 2 and 3.

For generating Bayesian designs, assumptions have to be made concerning the uncertainty of the prior estimate on top of determining β_0 . In this paper, we used a 6-dimensional normally distributed prior with mean $\beta_0 = [-0.5 \ 0 \ -0.5 \ 0 \ -0.5 \ -1]$ and variance-covariance matrix

$$\begin{bmatrix} 0.5 \cdot \mathbf{I}_{M-1} & \mathbf{0}_{(M-1) \times 1} \\ \mathbf{0}_{1 \times (M-1)} & 0.1 \end{bmatrix}, \quad (23)$$

where \mathbf{I}_{M-1} is the $(M - 1)$ -dimensional identity matrix, for expressing the prior belief about the model parameters and the corresponding uncertainty. The first five elements of the mean vector correspond to the utility coefficients associated with the three attributes. The last element corresponds to the negative utility coefficient related to the price. Assuming that the two-level attribute is coded as +1 for the first level and -1 for the second level, the prior implies that the utility of a commodity increases with the attribute levels, except for the price attribute. As can be seen in (23), the variance for the price coefficient is smaller than the variance for the other coefficients in order to ensure that only negative price coefficients are taken into account in the Bayesian optimal design approach. Furthermore, this prior distribution follows the recommendations formulated in Kessels et al. [2006b].

4.2 Evaluation criteria

We evaluate the designs based on two criteria: the ability to estimate the WTP accurately and the ability to yield accurately estimated utility coefficients. To compute these two measures, we estimate the utility coefficients of the conditional logit model based on simulated observations. These estimated coefficients are then used to compute the WTP . Comparing the estimated utility coefficients with their real counterparts and the estimated WTP with the real WTP gives us an idea of the estimation accuracy of the designs.

The two evaluation criteria used to assess the designs are the squared error between the real and estimated utility coefficients and the squared error between the real and the estimated WTP values and take the following form

$$EMSE_{\beta}(\beta) = \int_{\mathfrak{R}^M} (\hat{\beta} - \beta)' (\hat{\beta} - \beta) f(\hat{\beta}) d\hat{\beta} \quad (24)$$

and

$$EMSE_{WTP}(\beta) = \int_{\mathfrak{R}^M} (\widehat{\mathbf{W}}(\hat{\beta}) - \mathbf{W}(\beta))' (\widehat{\mathbf{W}}(\hat{\beta}) - \mathbf{W}(\beta)) f(\hat{\beta}) d\hat{\beta} \quad (25)$$

where $\widehat{\mathbf{W}}$ and \mathbf{W} are vectors containing the WTP estimates and the real WTP values, respectively, and $f(\hat{\beta})$ represents the distribution of the estimated utility coefficients. These evaluation criteria capture the bias and the variability of the estimates and are therefore frequently used measures. The expressions are approximated by generating 1 000 data sets for a given value of β and are calculated for 150 values of the parameter β drawn from a 6-dimensional normal distribution. A small value of $EMSE$ is preferred as this indicates accurate estimates.

Since the estimated price coefficient enters the WTP computation in a nonlinear way, a poorly estimated price coefficient can result in unrealistic WTP estimates and consequently in unreasonably high values of $EMSE_{WTP}$. The problem of unrealistic WTP estimates has already been described by several authors (among others Sonnier et al. [2007] and Scarpa et al. [Forthcoming]). To compare the designs, we compute some relevant simulation statistics of the distribution of the 150 $EMSE_{WTP}$ or $EMSE_{\beta}$ values: the average, the median, the standard deviation, the min-

Simulation statistics of the distribution of the $EMSE_{WTP}$ values							
Design	Average	Median	St.dev.	Min	5% perc.	95% perc.	Max
D^{pref} -opt.	2.94E+02	0.0678	3.21E+03	0.0158	0.0223	12.4971	3.90E+04
A^{pref} -opt.	7.38E+02	0.0925	8.76E+03	0.0150	0.0196	1.5281	1.07E+05
c^{pref} -opt.	2.70E+01	0.0958	3.28E+02	0.0377	0.0472	0.6939	4.02E+03
$minimax^{pref}$ opt.	4.34E-01	0.1144	1.65E+00	0.0424	0.0551	1.1119	1.77E+01
$varmin^{pref}$ opt.	2.22E-01	0.1141	4.70E-01	0.0458	0.0634	0.5913	3.80E+00
fid^{pref} opt.	2.52E-01	0.1060	6.22E-01	0.0521	0.0670	0.6362	5.34E+00
M^{pref} -opt.	4.44E+05	0.1000	5.44E+06	0.0161	0.0241	17.9287	6.66E+07
D^{WTP} -opt.	4.35E-01	0.1041	1.80E+00	0.0261	0.0394	1.7513	2.14E+01
A^{WTP} -opt.	2.41E-01	0.0971	4.67E-01	0.0272	0.0398	0.8168	3.78E+00
$varmin^{WTP}$ opt.	6.18E-01	0.1087	3.61E+00	0.0256	0.0356	1.0957	4.24E+01

Table 2: Simulation statistics of the distribution of $EMSE_{WTP}$ values yielded by the different designs assuming correct prior information

imum, the 5% and 95% percentile and the maximum of the distribution are useful statistics to represent the spread of the estimation errors.

4.3 Performance of the designs assuming correct prior information

In this section, we study the ability of the designs to yield precise utility coefficient and WTP estimates when the prior information is correct. This means that the 150 true parameter vectors β are drawn from the prior distribution used to generate the designs. We assume 50 respondents taking part in the experiment.

The WTP estimation accuracy of the different designs is represented in Table 2 by means of some statistics on the $EMSE_{WTP}$ values. It can be seen from this table that the $fiducial^{pref}$ optimal, the $varmin^{pref}$ optimal and the A^{WTP} -optimal design estimate the WTP most accurately as the averages and the medians of their distributions of the $EMSE_{WTP}$ values are lowest. This shows that they are most successful in reducing the size and the number of outlying WTP estimation errors. Although the other designs have a median which is more or less comparable in size, they all result in a higher average of the distribution because of more and larger outlying values. The D^{pref} -optimal, the A^{pref} -optimal, the c^{pref} -optimal and the M^{pref} -optimal design show the worst results in this respect: their average WTP estimation errors are highest and their maximum WTP

Simulation statistics of the distribution of the $EMSE_{\beta}$ values							
Designs	Average	Median	St.dev.	Min	5% perc.	95% perc.	Max
D^{pref} -opt.	0.1076	0.0843	0.1367	0.0411	0.0482	0.1851	1.4388
A^{pref} -opt.	0.1213	0.0788	0.3087	0.0413	0.0482	0.1674	3.1840
c^{pref} -opt.	0.1175	0.1118	0.0348	0.0565	0.0706	0.1824	0.2157
$minimax^{pref}$ opt.	0.2283	0.1316	0.5574	0.0644	0.0795	0.5937	6.4238
$varmin^{pref}$ opt.	0.1475	0.1326	0.0836	0.0641	0.0829	0.2361	0.9875
fid^{pref} opt.	0.1707	0.1254	0.2617	0.0664	0.0827	0.3447	3.1884
M^{pref} -opt.	0.1156	0.0979	0.0983	0.0464	0.0542	0.2357	1.1299
D^{WTP} -opt.	0.2596	0.1226	0.5082	0.0452	0.0601	0.9936	3.7817
A^{WTP} -opt.	0.1890	0.1130	0.3133	0.0517	0.0635	0.4783	3.2222
$varmin^{WTP}$ opt.	0.1236	0.1059	0.0905	0.0584	0.0685	0.2090	0.8508

Table 3: Simulation statistics of the distribution of the $EMSE_{\beta}$ values yielded by the different designs assuming correct prior information

estimation errors are much larger than those of the other design options.

Table 3 contains the statistics on the $EMSE_{\beta}$ values which give the estimation accuracy of the utility coefficients obtained by the different designs. As expected, the D^{pref} - and A^{pref} -optimal design estimate the utility coefficients most precisely as these designs yield the lowest median of the distribution of the estimation errors. It can be seen that the M^{pref} -optimal design estimates the utility coefficients almost as accurately as the D^{pref} -optimal and the A^{pref} -optimal design. These designs are closely followed by the c^{pref} -optimal and the $varmin^{WTP}$ -optimal design. These four designs exhibit as well the lowest average estimation error of the utility coefficients. Furthermore, Table 3 indicates that the enhanced ability to estimate the WTP accurately comes at a small cost in terms of the estimation precision of the utility coefficients.

4.4 Performance of the designs assuming incorrect prior information

So far, the assumption was made that the prior distribution on β contains correct information on the utility coefficients. In this section, we relax this assumption and study the performance of the designs in case of wrongly specified prior information. The same designs as in Section 4.3 were used to generate the data, but the distribution used to draw the true parameters differs from the prior distribution used to generate the designs. We assume that 50 respondents participate

Simulation statistics of the distribution of the $EMSE_{WTP}$ values							
Designs	Average	Median	St.dev.	Min	5% perc.	95% perc.	Max
D^{pref} -opt.	1.42E+05	2.7000	1.46E+06	0.1872	0.4540	1.07E+04	1.75E+07
A^{pref} -opt.	1.23E+04	3.5200	1.19E+05	0.1955	0.4799	1.46E+04	1.45E+06
c^{pref} -opt.	1.76E+02	1.2292	1.37E+03	0.2029	0.3536	3.75E+01	1.46E+04
$minimax^{pref}$ opt.	8.27E+00	1.8168	1.41E+01	0.2769	0.4755	4.24E+01	6.45E+01
$varmin^{pref}$ opt.	4.99E+00	1.1848	9.15E+00	0.2589	0.4585	2.65E+01	5.41E+01
fid^{pref} opt	4.20E+00	1.1613	7.96E+00	0.2197	0.4280	2.13E+01	5.38E+01
M^{pref} -opt.	1.78E+02	2.1325	1.37E+03	0.0764	0.3976	1.33E+02	1.59E+04
D^{WTP} -opt.	2.15E+04	2.1800	2.13E+05	0.0615	0.3418	3.31E+03	2.57E+06
A^{WTP} -opt.	7.01E+00	1.5462	1.86E+01	0.2379	0.4475	3.34E+01	1.88E+02
$varmin^{WTP}$ opt.	7.84E+05	3.9000	9.60E+06	0.2980	0.7737	2.63E+02	1.18E+08

Table 4: Simulation statistics of the distribution of the $EMSE_{WTP}$ values yielded by the different designs assuming incorrect prior information

in the experiment. In a first scenario, the true parameters β come from a 6-dimensional normal distribution with mean $[-1 \ 0 \ -1 \ 0 \ -1 \ -0.5]$ and variance-covariance matrix

$$\begin{bmatrix} \mathbf{I}_{M-1} & \mathbf{0}_{(M-1) \times 1} \\ \mathbf{0}_{1 \times (M-1)} & 0.0025 \end{bmatrix}. \quad (26)$$

This implies that the prior underestimated most effects and does not cover all true parameter vectors.

Table 4 shows the statistics on the $EMSE_{WTP}$ values. It indicates that the $fiducial^{pref}$ optimal design estimates the WTP the most accurately: it exhibits the lowest average, median, standard deviation and maximum value of the WTP estimation errors. The $varmin^{pref}$ optimal and A^{WTP} -optimal design perform only a little worse than the $fiducial^{pref}$ optimal design. Compared to these three designs, the c^{pref} -optimal design has a higher average and median value. Its standard deviation is larger and the design is less successful in reducing the size of the outlying values. All the other designs are clearly outperformed by the $fiducial^{pref}$ optimal design, the $varmin^{pref}$ optimal design and the A^{WTP} -optimal design.

Table 5 lists the statistics on the $EMSE_{\beta}$ values using the different designs and assuming incorrect prior information. Based on the median and the average of the estimation errors, it can be seen

Simulation statistics of the distribution of the $EMSE_{\beta}$ values							
Designs	Average	Median	St.dev.	Min	5% perc.	95% perc.	Max
D^{pref} -opt.	0.4254	0.1861	0.7401	0.0500	0.0689	1.8231	6.7664
A^{pref} -opt.	0.4622	0.1537	1.1227	0.0467	0.0636	2.3182	8.1819
c^{pref} -opt.	0.3986	0.1448	0.8281	0.0533	0.0694	1.9917	6.3500
$minimax^{pref}$ opt.	1.6493	0.2543	2.9409	0.0724	0.0939	8.4852	14.4558
$varmin^{pref}$ opt.	0.8469	0.1726	1.7113	0.0763	0.0918	5.3933	9.4517
fid^{pref} opt.	0.8182	0.2117	1.6384	0.0742	0.0961	4.8621	9.7954
M^{pref} -opt.	0.3752	0.1645	1.1950	0.0467	0.0712	0.8263	14.3829
D^{WTP} -opt.	0.3763	0.1779	0.7380	0.0495	0.0647	1.5762	6.7147
A^{WTP} -opt.	0.9434	0.1971	1.9403	0.0658	0.0839	5.9907	11.7927
$varmin^{WTP}$ opt.	1.3732	0.2206	3.0451	0.0727	0.0957	7.8439	17.5498

Table 5: Simulation statistics of the distribution of the $EMSE_{\beta}$ values yielded by the different designs assuming incorrect prior information

that the D^{pref} -optimal, the A^{pref} -optimal, the c^{pref} -optimal, the M^{pref} -optimal and the D^{WTP} -optimal design estimate the utility coefficients the most precisely. The more accurate estimation of the WTP comes at a small cost in terms of the estimation precision of the utility coefficients. However, the difference is not sufficiently large to rule out one of the designs.

In a second scenario, the real utility coefficients used in the data generation process come from a 6-dimensional normal distribution with mean $[0\ 0\ 0\ 0\ 0\ -0.5]$ and variance-covariance matrix

$$\begin{bmatrix} 0.25 \cdot \mathbf{I}_{M-1} & \mathbf{0}_{(M-1) \times 1} \\ \mathbf{0}_{1 \times (M-1)} & 0.0025 \end{bmatrix}. \quad (27)$$

This implies that individuals are almost indifferent between the different attribute levels of a commodity, except the price. Results for this scenario are not shown as the conclusions are similar to those of the first scenario.

This simulation study leads to the conclusion that *fiducial*^{pref} optimal designs, the *varmin*^{pref} optimal designs and the A^{WTP} -optimal designs estimate the WTP accurately no matter whether the prior information to create the design is correct or not. Moreover, the designs generated by these criteria seem to reduce most successfully the number and the size of extreme WTP estimates. The disadvantage of these designs is a slightly reduced performance in estimating the

utility coefficients. Although the *fiducial^{pref}* optimal designs lead to the best results, they are the hardest of the three designs to compute.

5 Illustration

Hole [2008] examines the patient’s preference for appointments with a general practitioner (GP) by means of a discrete choice experiment. By including the cost for the patient, the estimated utility coefficients of the conditional logit model can be used to calculate his/her willingness-to-pay for these attributes. In this way, the government is able to align health services to the preferences of the population and to quantify to which extent it is prepared to pay for quality. In this section, we examine the performance of the designs proposed in previous sections and compare them with the design strategy used in the original study.

The experiment investigated the *WTP* for five attributes of an appointment with a GP: waiting time for an appointment, flexibility of the time of the appointment, the doctor’s interpersonal manner, the doctor’s knowledge of the patient and the thoroughness of the examination. To be able to compute the *WTP* for the different attributes, the patient’s cost for an appointment is considered. The levels of the attributes are given in Table 6. All attributes are dummy coded, except the waiting time and the cost for the patient which are linearly coded.

Attribute	Levels
Waiting time for appointment	same day, next day, 2 or 5 days
Cost for the patient	£0, £8, £18 or £28
Doctor’s knowledge of patient	knows you well or not
Flexibility of time	one appointment offered or choice
Doctor’s interpersonal manner	friendly and warm or formal and businesslike
Thoroughness of examination	thorough or not

Table 6: Attributes of the experiment eliciting a patient’s preference for a GP appointment

Each of the 409 respondents had to choose his or her preferred alternative in eight choice sets of size two. The original study used a locally *D*-optimal design with the attribute coefficients set to zero which assumes that the patient is indifferent between the attribute levels. The estimated

utility coefficients, their t-statistics and corresponding WTP estimates obtained after analysis of the choices by a conditional logit model are shown in Table 7.

Attribute	$\hat{\beta}$	t-stat.	\widehat{WTP}
Waiting time	-0.131	-7.71	-1.71
Cost for patient	-0.077	-27.22	/
Dr's knowledge	0.344	6.49	4.48
Flexibility of time	0.194	3.60	2.53
Dr's interpers. manner	0.317	6.13	4.13
Thoroughness	1.061	18.24	13.82
Constant	-0.023	-0.40	/

Table 7: Parameter estimates, t-statistics and WTP resulting from the original GP experiment

We compare the Bayesian c^{pref} -optimal, the Bayesian $fiducial^{pref}$ optimal, the Bayesian $varmin^{pref}$ optimal, the Bayesian A^{WTP} -optimal design, the locally and Bayesian D^{pref} -optimal design in terms of their ability to yield precise WTP estimates. We exclude the A^{pref} -optimal, M^{pref} -optimal, $minimax^{pref}$ optimal, D^{WTP} -optimal and $varmin^{WTP}$ optimal designs in this evaluation as they were outperformed by the others in the simulation study in Section 4. We include the locally D^{pref} -optimal design generated assuming zero prior coefficients as this design was used in Hole [2008]. To generate the Bayesian optimal designs under consideration, we assume a 7-dimensional normally distributed prior with mean $\mathbf{0}_{1 \times 7}$ and variance $0.25 \cdot \mathbf{I}_7$.

Choices for 409 participants are generated using the utility coefficient estimates of the original study that are displayed in Table 7. The WTP estimates for the attributes are computed from the estimated utility coefficients of the conditional logit model. This simulation study is performed for 2 000 data sets. We assess the different designs by examining the distribution of the 2 000 WTP estimates for each attribute.

Table 8 shows the statistics of the distribution of the WTP estimates for the attribute 'Interpersonal manner'. We do not present the results for the other attributes as they lead to similar conclusions. It can be seen that the Bayesian D^{pref} -optimal design does not yield accurate WTP estimates. This design also results in the largest outlying WTP values. The estimation precision obtained by

Distribution of the WTP estimates for 'Interpersonal manner'							
Designs	Average	Median	St.dev.	Min	5 perc.	95 perc.	Max
Loc. D^{pref} -opt.	4.2072	4.1588	0.7365	1.6840	3.1469	5.4133	9.4130
Bays. D^{pref} -opt.	7.1533	4.0558	160.9592	-3422.4283	2.2978	12.6046	6256.5390
Bays. c^{pref} -opt.	4.1424	4.1511	0.9021	1.4307	2.6192	5.6272	7.1413
Bays. fid^{pref} opt.	4.1366	4.0996	0.4975	2.7231	3.3464	5.0062	6.0905
Bays. $varmin^{pref}$ opt.	4.1021	4.1398	1.0960	0.1857	2.3149	5.9347	8.6526
Bays. A^{WTP} -opt.	4.0935	4.0925	1.0734	0.2226	2.3075	5.8136	7.1806

Table 8: Statistics of the distribution of the WTP estimates for the attribute 'Interpersonal manner' for the different designs when $\widehat{WTP}_{int.man.} = 4.13$

Distribution of the WTP estimates for 'Interpersonal manner'							
Designs	Average	Median	St.dev.	Min	5 perc.	95 perc.	Max
Loc. D^{pref} -opt.	8.3441	8.4905	1.4816	4.8260	6.0196	10.6010	16.3688
Bays. D^{pref} -opt.	14.4053	6.9037	203.3741	-1480.2320	-18.3102	33.6259	6915.8390
Bays. c^{pref} -opt.	7.4350	7.4668	1.3791	1.8227	5.1240	9.6403	12.1934
Bays. fid^{pref} opt.	7.5114	7.4743	0.7227	5.4147	6.3557	8.7379	9.9137
Bays. $varmin^{pref}$ opt.	7.6152	7.6229	1.0330	4.9169	5.9484	9.3456	11.7827
Bays. A^{WTP} -opt.	7.5227	7.5459	0.8837	4.4050	6.0675	8.9688	10.4274

Table 9: Statistics of the distribution of the WTP estimates for the attribute 'Interpersonal manner' for the different designs when $\widehat{WTP}_{int.man.} = 7.50$

the other designs does not differ sufficiently to rule out one of them.

However, it has to be noticed that the zero prior utility coefficients used to develop the locally D^{pref} -optimal design are close to the utility coefficients used to generate the data. This gives the locally D^{pref} -optimal design an advantage when comparing with the Bayesian designs which take the uncertainty on the prior formally into account. That is why we also study the case in which the data are generated with coefficients $[-0.75 \ -0.10 \ 0.75 \ 0.75 \ 0.75 \ 0.75 \ 0]$. The statistics of the distribution for the WTP estimates for the attribute 'Interpersonal manner' are given in Table 9. Only the results for this attribute are given as the conclusions for the other attributes are similar. It can be seen in Table 9 that the locally D^{pref} -optimal design is not robust against incorrect prior information as it leads to less accurate WTP estimates and larger outlying values than the Bayesian design criteria, except the Bayesian D^{pref} -optimal one. The *fiducial*^{pref} optimal design yields

the most accurate WTP estimates and reduces the size of the largest WTP most successfully.

This illustration confirms that the $fiducial^{pref}$ optimality criterion results in precise estimates of the WTP whether the prior information used to develop the design is correct or not. However, in this illustration, the difference between the other proposed design efficiency criteria to estimate the WTP precisely is not sufficiently large to exclude one of them. Although, it is clear that more precise WTP estimates can be obtained by accommodated design criteria than by the D^{pref} -optimality criterion.

6 Conclusion

To estimate the WTP accurately, Vermeulen et al. [2008] applied a c^{pref} -optimality criterion to create conjoint choice experiments. Although, this criterion yields more accurate WTP estimates than several benchmark designs including D -optimal, orthogonal, random and balanced overlap designs, it still leads to extreme estimates. In this paper, other optimality criteria which reduce the occurrence of extreme WTP estimates as much as possible are considered.

We distinguish between two categories of criteria: (i) criteria in preference space, which specifies the utility of a commodity in terms of the utility coefficients associated with the attribute levels, and (ii) criteria in WTP -space, which defines the utility in terms of the WTP and the price coefficient. We show how to create Bayesian optimal designs in these two spaces and compare them in terms of the accuracy of the WTP estimates and estimated utility coefficients.

A simulation study and a numerical example reveal that the $fiducial^{pref}$ optimal, the $varmin^{pref}$ optimal and the A^{WTP} -optimal designs allow accurate WTP estimates. This result is valid no matter whether the prior information on the utility coefficients is correct or not. Moreover, these designs are more successful in reducing the occurrence of extremely large estimates of the WTP than all other design options. The enhanced ability to estimate the WTP accurately comes only at a small cost in estimation accuracy of the utility coefficients themselves. Although the $fiducial^{pref}$ optimal designs lead to the best results, they are the hardest of the three designs to compute.

What is clear though, is that using the D^{pref} -optimality or the A^{pref} -optimality criterion to create designs for the purpose of estimating the WTP yields less accurate WTP estimates and results in a large number of extreme WTP values, although they still perform better than standard designs including orthogonal, random and balanced overlap designs.

7 Acknowledgements

Bart Vermeulen is funded by project G.0611.05 of the Fund for Scientific Research Flanders.

References

- Alberini, A. (1995), "Optimal designs for discrete choice contingent valuation surveys: single-bound, double-bound, and bivariate models," *Journal of Environmental Economics and Management*, 28, 287–306.
- Banfi, S.; Farsi, M.; Filippini, M. and Jacob, M. (2008), "Willingness to pay for energy-saving measures in residential buildings," *Energy Economics*, 30, 503–516.
- Boxall, P. C. and Adamowicz, W. L. (2002), "Understanding heterogeneous preferences in random utility models: a latent class approach," *Environmental and Resource Economics*, 23, 421–446.
- Carlsson, F. and Martinsson, P. (2001), "Do hypothetical and actual marginal WTP differ in choice experiments?" *Journal of Environmental Economics and Management*, 41, 179–192.
- 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*, 53, 342–363.
- Hensher, D. and Sullivan, C. (2003), "Willingness to pay for road curviness and road type," *Transportation Research Part D*, 8, 139–155.
- Hole, A. (2008), "Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment," *Journal of Health Economics*.
- Hole, A. R. (2007), "A Comparison Of Approaches To Estimating Confidence Intervals For Willingness To Pay Measures," *Health Economics*, 16, 827–840.
- Kanninen, B. J. (1993), "Optimal Experimental Design for Double-Bounded Dichotomous Choice Contingent Valuation," *Land Economics*, 69, 138–146.
- Kessels, R.; Goos, P. and Vandebroek, M. (2006a), "A comparison of criteria to design efficient choice experiments," *Journal of Marketing Research*, 43, 409–419.
- Kessels, R.; Jones, B.; Goos, P. and Vandebroek, M. (2006b), "Recommendations on the use of Bayesian optimal designs for choice experiments," Research Report KBI_0617, Department of Decision Sciences and Information Management, Katholieke Universiteit Leuven, 18pp.

- Meyer, R. and Nachtsheim, C. (1995), “The coordinate-exchange algorithm for constructing exact optimal experimental designs,” *Technometrics*, 37, 60–69.
- Rose, J. M. and Scarpa, R. (2008), “Designs efficiency for nonmarket valuation with choice modelling: how to measure it, what to report and why,” *Australian Journal of Agricultural and Resource Economics*.
- Ryan, M. (2004), “A comparison of stated preference methods for estimating monetary values,” *Health Economics*, 13, 291–296.
- Sammer, K. and Wüstenhagen, R. (2006), “The influence of eco-labelling on consumer behaviour - results of a discrete choice analysis for washing machines,” *Business Strategy and Environment*, 15, 185–199.
- Sándor, Z. and Wedel, M. (2001), “Designing Conjoint Choice Experiments Using Managers’ Prior Beliefs,” *Journal of Marketing Research*, 38, 430–444.
- (2005), “Heterogeneous conjoint choice designs,” *Journal of Marketing Research*, 42, 210–218.
- Scarpa, R.; Thiene, M. and Train, K. (Forthcoming), “Utility in WTP space: a tool to address confounding random scale effects in destination choice to the Alps,” *American Journal of Agricultural Economics*.
- Sonnier, G.; Ainslie, A. and Otter, T. (2007), “Heterogeneity distributions of willingness-to-pay in choice models,” *Quantitative Marketing and Economics*, 5, 313–331.
- Toubia, O. and Hauser, J. (2007), “On managerial efficient experimental designs,” *Marketing Science*, 26, 851–858.
- Train, K. (2003), *Discrete Choice Methods with Simulation*, New York: Cambridge University Press.
- Train, K. and Weeks, M. (2005), “discrete choice models in preference space and willingness-to-pay space,” in *Applications of simulation methods in environmental and resource economics*, eds. Scarpa, R. and Alberini, A., Dordrecht, The Netherlands: Springer Publisher.

Vermeulen, B.; Goos, P.; Scarpa, R. and Vandebroek, M. (2008), “Efficient and robust willingness-to-pay designs for choice experiments: some evidence from simulations,” Research Report KBI_0809, Department of Decision Sciences and Information Management, Katholieke Universiteit Leuven, 18pp.

Wang, Y. (2000), “Fiducial intervals: what are they?” *The American Statistician*, 54, 105–111.