G. Stefani[*] and R. Scarpa[†]

Solving the flip of the coin/unidentification dilemma : a Bayesian perspective on the incentive compatibility of hypothetical referenda

Abstract

We propose a Bayesian approach that overcomes the identification of scale in referendum contingent valuation data with experimental treatment. We apply this approach to the data originally collected by Cummings *et al.* (1997) and commented by Haab *et al.* (1999) and further rejoined by other studies. The results support the substantive findings of a higher rate of "yes" responses under hypothetical treatment, and provide a feasible solution for the lack of identification of the scale parameter.

Keywords: Contingent Valuation, Incentive compatibility, scale identification, discrete choice modelling, Bayesian modelling

JEL classification: C11, C25, Q50

Acknowledgments: The authors would like to thank, with the usual disclaimer, Tim Haab, Glenn Harrison, Fredrik Carlsson and Federico Mattia Stefanini for the comments provided on various versions of this paper.

[*] Dipartimento di Scienze per l'Economia e l'Impresa – Dipartimento di Scienze delle Produzioni Vegetali, del Suolo e dell'Ambiente – Università di Firenze, Italy

[†]Department of Economics, Waikato Management School, University of Waikato, New Zealand

1. Introduction

We propose a Bayesian approach that overcomes the identification of scale in referendum contingent valuation data with experimental treatment. This issue has been raised in the literature following an initial paper in Cummings *et al.* (1995) where the authors reported results from an experiment devised to test the evidence of hypothetical bias in referendum contingent valuation for a private good. The rate of yes to no responses between hypothetical (84 percent) and real (27 percent) payment settings differed significantly from each other. However, as most of the CVM studies are carried out on public goods[‡], in a subsequent paper Cummings, Elliot, Harrison and Murphy (1997) (henceforth CEHM) tried to replicate their result also for this type of goods.

To test the incentive compatibility assumption for hypothetical referendum CEHM chose as a public good the support for an information campaign (a booklet to be distributed among citizens) about water contamination in New Mexico employing as payment vehicle a donation to a Charity. The public nature of the good stems from the fact that if the majority supports the proposal, then every participant in the referendum must pay for the provision of the good. CEHM found evidence in support of hypothetical bias also in the public good case. However, their paper soon was the object of two comments on the Journal of Political Economy. The first one by Smith (1999) was about the scope of the provisioned public goods, the second by Haab, Hang and Whithehead (1999) (henceforth HHW) criticised the method employed to test the incentive compatibility hypothesis. HHW claimed that CEHM results relied on the hypothesis of homoskedasticity between the real and hypothetical treatment. Failing to account for possible differences of variability of responses across treatments would have resulted in inconsistent parameter inference. In particular, once the higher variability of hypothetical responses is taken into account no significant difference appears between real and hypothetical response rate supporting the hypothesis of incentive compatibility.

[‡] Differences in incentive compatibility of hypothetical referenda between private and public goods have been suggested on theoretical grounds by Carson and Groves (2007) and reviewed in empirical studies by List and Gallet (2001) and by Little and Berrens (2004).

HHW also stress that in order to test for the presence of hypothetical bias the referendum data employed by CEHM are not suitable since the scale parameters cannot be uniquely estimated. To overcome this problem they suggest employing multiple bids designs to examine the incentive compatibility issue. In a paper by Carlsson and Johansson-Stenman (2010) (abbreviated in what follows as CJ-S) the appropriateness of the methodology used by HHW is questioned and the difficulties arising when one attempts to properly identify the scale parameter in referendum data are highlighted. CJ-S also address the broader issue of identification in this context revising the evidence stemming from the CEHM dataset.

In this context we propose a Bayesian estimator for the heteroskedastic probit, based on plausible priors, could mitigate the problems encountered by maximum likelihood estimators thereby providing an alternative tool for the analysis of this type of data. The rest of the note is set out as follows: in the second section we review the methodological issues raised by the CEHM paper. In the next section a Bayesian version of the CEHM model is illustrated and the results are discussed. Finally, in the last section some conclusions are drawn.

2. The CEHM model: methodological issues

In CEHM the hypothesis of absence of hypothetical bias is tested by estimating a probit model where the probability of observing a "yes" vote is explained by some socio-demographics and a dummy treatment variable (real vs. hypothetical). The results indicate that the only significant coefficient in the regression is the treatment variable either when considering the effect on the latent variable or the marginal effect of treatment on such probability (evaluated at sample means).

In their comment HHW argue that the parameter estimate of the treatment variable is inconsistent because CEHM fail to account for the higher variability of responses in the hypothetical setting where the opportunity cost of deviating from rational responses is lower. Following Cameron and James (1987) HHW assume that WTP for the good is the latent variable behind the probit model, then if the WTP is specified as a linear function of socio-demographic covariates (**X**) and a normally distributed random error term (ε) the probability of a no response is given by:

$$\Pr(WTP_i < t) = \Pr\left(\mathbf{X}_i\boldsymbol{\beta} + u_i < t\right) = P\left(\frac{u_i}{\sigma} < \frac{t - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right) = \Phi\left(\frac{t - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right) \qquad (1)$$

where σ is the standard deviation of the error term and $\boldsymbol{\beta}$ is a vector of parameters for the linear function. As t is not varied across individuals (differently from most CVM dichotomous choice studies) it is absorbed into the constant (β_0). According to HHW, ignoring differences in both scale and parameters across treatments and checking for significance of a treatment dummy within a pooled regression as CEHM did, results in testing the following hypothesis:

$$H_0 : \frac{\beta_{0,r}}{\sigma_r} = \frac{\beta_{0,h}}{\sigma_h} \text{ instead of } H_0 : \beta_{0,r} = \beta_{0,h} \tag{2}$$

while maintaining equality of normalised parameters for other variables ($\boldsymbol{\beta}_{X,r}/\sigma_r = \boldsymbol{\beta}_{X,h}/\sigma_h$) between the two treatments.

In order to properly analyse the data HHW follow a procedure suggested by Swait and Louviere (1993) to test the joint and separate hypothesis of scale and parameter equality. First, the relative scale parameter $\mu = \sigma_r/\sigma_h$ is estimated through a grid search procedure over possible values of μ in order to maximise the likelihood of the probit on the data matrix:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_h \\ \mu\mathbf{X}_r \end{pmatrix} \tag{3}$$

Once a value for μ is obtained, estimation under the hypothesis of equality of beta parameters across treatments is achieved by estimating a probit on the data matrix where the data for the real treatment are scaled by the likelihood-maximizing μ*. The alternative of inequality of parameters with no restriction on scale is estimated by running independent probits on the public and private treatment data, and finally a LR test is performed. If the hypothesis of equality of parameters is not rejected then the hypothesis of equality of variances can be tested by comparing the original pooled model with the rescaled pooled model. HHW find a relative scale parameter of 0.0406, and fail to reject the hypothesis of equality of parameters although the hypothesis of equality of variances was strongly rejected. Their result implies that the variance for the hypothetical subsample is about 600 times larger than for the real subsample: a very high value, which we argue to be empirically implausible when the context of choice is $10 for the provision of two citizen guides on the use of contaminated groundwater.

CJ-S criticise the findings by HHW, arguing that the procedure they adopt is not the one originally proposed by Swait and Louviere (1993). The key point made by CJ-S is that the estimation of the relative scale parameter performed by HHW omitted the treatment variable from the dataset. This results in violates the warning that the procedure will work only under a true model that includes all appropriate variables in the WTP function (Swait and Louviere, 1993, n.1 at page 307). CJ-S try to re-estimate the scale factor under the appropriate specification (i.e. including the treatment variable) and obtain a monotonic likelihood function increasing in the value of μ, a finding that we can replicate. This result is due to the very weak information content in the data, which makes it impossible to estimate the scale parameter since the only significant parameter of the CEHM model is the treatment variable, affecting both the WTP equation *and* the scale. The conclusions by CJ-S (p. 9) are quite discouraging, as the invariance of the bid variable, together with that of other significant explanatory variables prevent the analyst from simultaneously identifying both the relative scale parameters and the shift variable.

However, in a recent reply to HHW, (Harrison, 2006a, 2006b) it is argued that it is not the main effect on the latent variable that matters; rather it is the marginal effect on the probability to vote "no" in the referendum that is of relevance. Indeed, the latter "takes into account the joint effect of the experimental treatment variable on the mean response and on the residual variance". In a classical heteroskedastic probit model the marginal effect of a variable included both in the latent variable and in the variance equation is given:

$$\frac{\partial\left(P(WTP<t)\right)}{\partial x_T} == \phi\left(\frac{t-\mathbf{x}\boldsymbol{\beta}}{\exp\left(\mathbf{w}'\boldsymbol{\gamma}\right)}\right)\frac{\beta_T-\left(\mathbf{x}'\boldsymbol{\beta}\right)\gamma_k}{\exp\left(\mathbf{w}'\boldsymbol{\gamma}\right)}, \qquad (4)$$

as shown in Greene(2004).

Results of a heteroskedastic probit on the CEHM dataset are reported in an unpublished study (Harrison, 2008) . The finding is that if group-wise heteroskedasticity for real and hypothetical treatments is taken into account then the parameter for the treatment dummy is not significant in the latent variable equation, yet the marginal effect on the probability to vote "yes" remains significant. This result contradicts the findings by HHW, who simulate response for each of the 275 individuals of the CEHM dataset employing 1,000 bootstrapped parameter vectors obtained from the split real and hypothetical samples. They find that the probability of a no response for each individual is the same for the real and hypothetical referenda.

However, when we reproduce this estimation in Nlogit (Greene, 2002), with the dummy variable for the treatment included in both the WTP and variance equations, we find the coefficient estimate for such a variable in the latent WTP equation to be abnormally large, suggesting some problem in the convergence process of the max likelihood algorithm, probably due to the lack of identification emphasised by CJ-S.

CEHM is certainly one of the studies in the environmental economics literature where the principle of replication of results has been duly applied. Analyses on those data have been carried out by several scholars, but the reliability and validity of the CEHM experiment is still questioned (and questionable) after ten years. We propose another approach to the analysis of the CEHM data under the Bayesian paradigm, which allows one to overcome the identification issue at the cost of invoking some mild prior assumptions.

3. A Bayesian Analysis of the CEHM model

Bayesian analysis of binary response data is particularly well equipped to deal with the sort of shortcomings highlighted by commentators of the CEHM study. In particular, the issue of identification in a Bayesian framework is less binding than in the classical approach since there is no need to impose exact restrictions to identify the model, rather, prior information is introduced by means of priors' distributions (see for example: Leamer 1978 ; Zellner ,1971). By assuming proper priors a posterior can be defined for the unidentified parameter space from which one can sample using a Monte Carlo Markov chain (MCMC) (see Rossy, Allemby and McCulloch, 2005 for details). Afterwards one may derive the posterior distribution for identified functions of the unidentified parameters. The technique of augmenting the data by simulating the latent variable underlying the probit model allows us to sample the posterior distribution of sigmas conditional on the betas in a straightforward manner.

The rest of this section is set out as follows: first the latent variable model is illustrated. Then the Gibbs' sampler is introduced together with the prior and posterior distributions. A further subsection is devoted to the issue of identification. Then we describe the algorithm we use and test it on simulated data. The last two subsections deal with the CEHM dataset and the application of the Bayesian model to it.

Data augmentation

We model the decision to vote "yes" in the referendum through a latent variable approach that accounts for groupwise heteroskedasticity in the data. Partitioning

Cumming's sample into the real (n_r=95) and the hypothetical (n_h= 178) subsamples[§] a stacked latent variable regression is obtained:

$$\begin{bmatrix} \mathbf{z}_r^* \\ \mathbf{z}_h^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}_r \\ \mathbf{X}_h \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{u}_r \\ \mathbf{u}_h \end{bmatrix} \tag{5}$$

where $\mathbf{z}_r^*$ and $\mathbf{z}_h^*$, are latent variables, $\mathbf{X}_r$ and $\mathbf{X}_h$ are $n_i \times k$ (with $i=r, h$) matrices of k explanatory variables (including a dummy variable for the treatment), and $\mathbf{u}_r$ and $\mathbf{u}_h$ are normally distributed errors with mean 0 and covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_h^2 \end{bmatrix} \tag{6}$$

The observational equivalent of the latent variable is defined as follows:

$$y_i = \begin{cases} = 0 & z_i < 0 \\ = 1 & z_i \geq 0 \end{cases} \tag{7}$$

where $y_i = 0$ and $y_i = 1$ respectively refer to "no" or "yes" votes to the referendum. The unknown parameters in the model are $\boldsymbol{\beta}, \sigma_r, \sigma_h$ together with the latent data $\mathbf{y}^* = \begin{bmatrix} \mathbf{z}_r^* & \mathbf{z}_h^* \end{bmatrix}'$.

Gibbs' sampler

We use a Gibbs' sampler with data augmentation to fit this model [1]. The method is standard for probit models, the only uncommon feature being the groupwise heteroskedasticity. With the Gibbs' sampler we simulate from a block conditional distribution of a θ vector of parameters partitioned into four sub-components or blocks $\boldsymbol{\theta} = \mathbf{y}^*, \boldsymbol{\beta}, \sigma_r, \sigma_h$. This method is commonly employed when it is difficult to sample directly from the marginal posterior density, but it is easy to sample from the conditional distributions for the individual blocks of parameters $\pi(\theta_k \mid \{\theta_j, j \neq k\})$.

[§] We discarded two observations from the original CEHM database as they show a missed value for the variable "race". Similar result are obtained setting to 1 the value for "race" (that is assuming the answer was "Caucasian") and employing the whole dataset.

Starting with initial guesses for the parameters $\left(\mathbf{z}^{(0)}, \boldsymbol{\beta}^{(0)}, \sigma_r^{(0)}, \sigma_h^{(0)}\right)$ the simulation goes on cyclically as follows, drawing in turn:

$z_i^{(1)}$ from $\pi\left(z_i \mid \boldsymbol{\beta}^{(0)}, \sigma_r^{(0)}, \sigma_h^{(0)}, y_i\right)$ (8)

$\boldsymbol{\beta}^{(1)}$ from $\pi\left(\boldsymbol{\beta} \mid \mathbf{z}^{(1)}, \sigma_r^{(0)}, \sigma_h^{(0)}, \mathbf{y}\right)$ (9)

$\sigma_r^{(1)}$ from $\pi\left(\sigma_r \mid \mathbf{z}^{(1)}, \boldsymbol{\beta}_r^{(1)}, \sigma_h^{(0)}, \mathbf{y}\right)$ (10)

$\sigma_h^{(1)}$ from $\pi\left(\sigma_r \mid \mathbf{z}^{(1)}, \boldsymbol{\beta}_r^{(1)}, \sigma_r^{(1)}, \mathbf{y}\right)$ (11)

and reiterating t steps. For t approaching infinity, the distribution of sampled values $\left\{\left(\mathbf{z}^{(t)}, \boldsymbol{\beta}^{(t)}, \sigma_r^{(t)}, \sigma_h^{(t)}\right)\right\}$ no longer depends on the initial guesses and it approximates the posterior distribution $\pi(\mathbf{z}, \boldsymbol{\beta}, \sigma_r, \sigma_h \mid \mathbf{y})$. So, if the simulation is run for t steps and the first t_0 draws are dropped to reduce the effect of the initial values, the remaining t-t_0 draws can be used to estimate posterior moments, the posterior probability density function, and other features of interest.

Priors and Posterior conditional distributions

Partitioning the n observation in $n = n_r + n_h$, with subscripts referring to the real and hypothetical sub-sample respectively, and assuming independence across observations, the likelihood function is given by:

$$L(\boldsymbol{\beta}, \sigma_r, \sigma_h) = \prod_{jr=1}^{nr} \Phi\left(\frac{\mathbf{x}_{jr}\boldsymbol{\beta}}{\sigma_r}\right)^{y_{jr}} \left[1 - \Phi\left(\frac{\mathbf{x}_{jr}\boldsymbol{\beta}}{\sigma_r}\right)\right]^{1-y_{jr}} \cdot \prod_{jh=1}^{nh} \Phi\left(\frac{\mathbf{x}_{jh}\boldsymbol{\beta}}{\sigma_h}\right)^{y_{jh}} \left[1 - \Phi\left(\frac{\mathbf{x}_{jh}\boldsymbol{\beta}}{\sigma_h}\right)\right]^{1-y_{jh}} \quad (12)$$

We adopted the following prior distributions for $\boldsymbol{\beta}, \sigma_r, \sigma_h$:

$$\beta \sim N_k\left(\boldsymbol{\mu}_\beta, \mathbf{V}_\beta\right) \quad (13)$$

$$\sigma_r \sim IG(a_r, b_r) \qquad , \ \sigma_h \sim IG(a_h, b_h) \quad (14)$$

With $\mathbf{V}_\beta = \mathbf{I} \cdot 100$, a=3 and b=4, we assume relatively diffuse priors to avoid imposing too restrictive a set of priors. The inverted gamma prior distribution for sigma with parameters a (shape) =3 and b (scale) =4 has mode equal to 1 and variance equal to 4. This prior distribution is centred on 1 and skewed to the right. It allows a substantial density over a large range of ratios of scale parameters. As such it represents a plausible, yet relatively diffuse, prior. The priors on $\boldsymbol{\mu}_\beta$ are obtained from the work by CEHM. The following posterior conditional densities are then obtained:

$$\pi\left(z_i \mid \boldsymbol{\beta},\sigma_r,\sigma_h,y_i\right) \sim ind \begin{cases} TN_{(-\infty,0]}\left(\mathbf{X}_i^{'}\boldsymbol{\beta},\boldsymbol{\Sigma}\right) & if \quad y_i = 0 \\ TN_{[0,\infty)}\left(\mathbf{X}_i^{'}\boldsymbol{\beta},\boldsymbol{\Sigma}\right) & if \quad y_i = 1 \end{cases} \text{ for } y_i\text{=1, } n \tag{15}$$

$$\pi(\boldsymbol{\beta} \mid \mathbf{z},\sigma_r,\sigma_h,\mathbf{y}) \sim N_k\left(\mathbf{D}_\beta \mathbf{d}_\beta, \mathbf{D}_\beta\right) \tag{16}$$

with $\mathbf{D}_\beta = \left(\mathbf{X}^{'}\boldsymbol{\Sigma}^{-1}\mathbf{X} + \mathbf{V}_\beta^{-1}\right)^{-1}$ and $\mathbf{d}_\beta = \mathbf{X}^{'}\boldsymbol{\Sigma}^{-1}\mathbf{y} + \mathbf{V}_\beta^{-1}\boldsymbol{\mu}_\beta$

$$\pi\left(\sigma_r \mid \mathbf{z},\sigma_h,\mathbf{y}\right) \sim IG\left(\frac{n_r}{2}+a_r, \left[b_r + \frac{1}{2}\left(\mathbf{y}_r - \mathbf{X}_r\boldsymbol{\beta}\right)^{'}\left(\mathbf{y}_r - \mathbf{X}_r\boldsymbol{\beta}\right)\right]\right) \tag{17}$$

$$\pi(\sigma_h \mid \mathbf{z},\sigma_r,\mathbf{y}) \sim IG\left(\frac{n_h}{2}+a_h, \left[b_h + \frac{1}{2}\left(\mathbf{y}_h - \mathbf{X}_h\boldsymbol{\beta}\right)^{'}\left(\mathbf{y}_h - \mathbf{X}_h\boldsymbol{\beta}\right)\right]\right) \tag{18}$$

where TN stands for truncated normal distribution.

Identification

The heteroskedastic probit model is still not identified in the stated form because the values of the dichotomous variable associated with the latent variable are identical if we the latter is multiplied by any arbitrary constant *c*, that is $\mathbf{y}(\mathbf{z}) = \mathbf{y}(c\mathbf{z})$. Therefore, as noted in the earlier discussion of the HHW model, the scale parameter is unidentified and:

$$\begin{bmatrix} c\mathbf{z}_r^* \\ c\mathbf{z}_h^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}_r \\ \mathbf{X}_h \end{bmatrix} c\boldsymbol{\beta} + \begin{bmatrix} c\mathbf{u}_r \\ c\mathbf{u}_h \end{bmatrix} \text{ with parameters } \begin{bmatrix} c\boldsymbol{\beta} & c^2\boldsymbol{\Sigma} \end{bmatrix} \tag{19}$$

describes the same observed y_i values as the original model does. Following Rossi, Allemby and McCulloch (2005, p. 107) we choose to leave the model unidentified by assuming a prior on the full set of unidentified parameters and obtaining the identified parameters by normalisation with respect to σ_r:

$$\text{x}\, \tilde{\boldsymbol{\beta}} = \frac{\boldsymbol{\beta}}{\sigma_r} \qquad \tilde{\sigma}_h = \frac{\sigma_h}{\sigma_r} \tag{20}$$

Algorithm

The estimation algorithms is as follows:

<u>Step 1</u>: Select starting values for $\boldsymbol{\beta}$, σ_r, σ_h. We choose to use maximum likelihood probit estimates for $\boldsymbol{\beta}$ in order to provide starting values of reasonable magnitude, while we set the initial values of σ_r and σ_h to 1.

Step 2: Draw initial vector of latent data ($\mathbf{z}$). We use the conditional posterior distribution (19). To draw from the truncated normal we employed the inverse transform method (Koop, Poirier and Tobias, 2007 p. 157).

Step 3: Draw $\boldsymbol{\beta}$ from the multivariate normal in eq. (20). In order to retrieve $\mathbf{D}_{\beta}$ the data matrix $\mathbf{X}$ is scaled with the appropriate weighting matrix $\boldsymbol{\Sigma}^{1/2}$ (as in a generalised least square estimator).

Step 4: Draw σ_r and σ_r from the inverse gamma distributions eq. (21) and eq. (22) employing the original data matrix to retrieve the error vector $\mathbf{y}_h - \mathbf{X}_h\boldsymbol{\beta}$.

Simulated data

In order to test the algorithm we run it with simulated data and with a sample of the same size as that used by CEHM. The simulated data was generated as follows:

$$\begin{bmatrix} \mathbf{z}_1^* \\ \mathbf{z}_2^* \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \tag{21}$$

with

$$\mathbf{x}_1 \sim U[-5,5] \tag{22}$$

$$\mathbf{x}_2 \sim U[-3,4] \tag{23}$$

$$\beta = \begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \tag{24}$$

Figure 1 illustrates the effect of scaling on the draws of parameters from the Gibb's posterior simulator. The draws from the marginal posterior of the unidentified β_1 wander erratically through the parameter space and do not show any tendency to converge while, the identified quantity β_1/σ_1 converge approximately on the true value of 1 used to generate the data set.

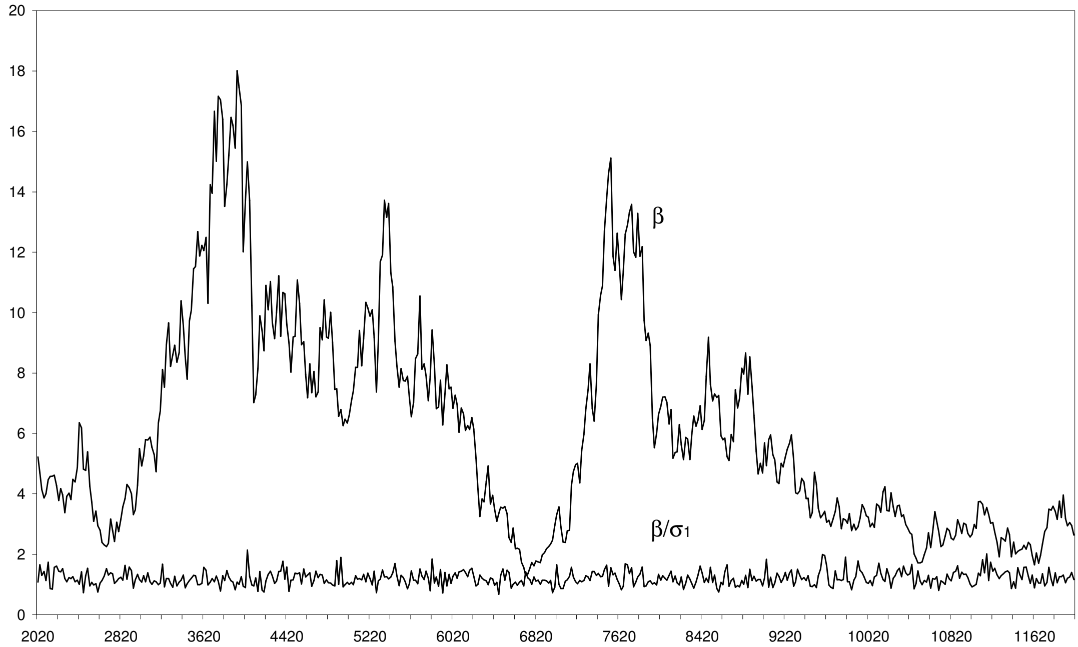

Figure 1: Gibbs sequence of β_1/σ_1 and unscaled β_1

Data

We employed the same dataset used by CEHM that was kindly provided by Harrison, who also provided us with the selection criteria for the sub-sample of experiments on which the original analysis was conducted.[**] For a description of the data we refer to the original paper by CEHM.

Results

Three independent Markov chains of length 30,000 were run to derive the posterior distributions of the scaled parameters. Starting values for $\boldsymbol{\beta}$ were randomly selected using perturbed maximum likelihood estimates of a homoskedastic probit model, while initial values for σ_r and σ_h were set to 1. An initial 10,000 burn-in replications were discarded to deal with the effect of the initial values while, to reduce correlation, a thinning factor of 20 was employed.

[**] We note that we use 273 rather than the 275 observations used by Haab, Huang and Whitehead(1999). The basic results, however, do not depend on these two missing observations.

Post estimation output analysis was performed with the CODA package (Plummer *et al.*, 2007) The potential scale reduction factor R (Gelman, Carlin and Rubin, 2004) is acceptable for all parameters in the model. We also run the Geweke (1992) convergence diagnostic test on each chain and CD values were lower than 1.96 for all parameters in the first chain, 7 and 11 out of 12 parameters in the second and third chain respectively[††]. The test for the parameter of the “real” dummy was always lower than 1.50. Summary statistics are illustrated in table 1. For all $\boldsymbol{\beta}$ parameters the value 0 is included within the 2.5 and the 97.5 quantiles of the posterior distribution.

Table 1. Descriptive statistics

	Gelman's R	Mean	SD	Naive SE	Time series SE	Q (0.025)	Q(0.50)	Q(0.975)
constant	1.00	-0.459	1.996	0.0364	0.0411	-4.877	-0.358	3.430
real	1.00	-0.492	0.261	0.0048	0.0048	-0.982	-0.502	0.019
rh	1.00	0.116	0.352	0.0064	0.0061	-0.638	0.129	0.774
age	1.00	0.012	0.010	0.0002	0.0002	-0.008	0.012	0.034
male	1.00	-0.111	0.205	0.0037	0.0045	-0.561	-0.091	0.243
race	1.00	-0.022	0.242	0.0044	0.0043	-0.468	-0.040	0.523
income	1.00	-0.002	0.006	0.0001	0.0001	-0.012	-0.002	0.011
married	1.00	0.133	0.224	0.0041	0.0040	-0.307	0.131	0.600
earn	1.01	-0.008	0.165	0.0030	0.0034	-0.341	-0.015	0.347
number	1.00	-0.005	0.023	0.0004	0.0005	-0.057	-0.003	0.038
student	1.00	0.158	0.340	0.0062	0.0070	-0.503	0.137	0.917
σ_h/σ_r	1.00	1.425	0.994	0.0182	0.0466	0.419	1.174	4.086

Note: Descriptive statistics are based on 36,000 iterations (3 independent chains of 12,000 iterations each)

However, we find mixed evidence that the coefficient for the "real" dummy is different from zero. Unlike the results obtained by HHW, who report a *t*-value of .39 in the rescaled probit, our results provide evidence in favour of some form of hypothetical bias, and hence in keeping with the original conclusions put forward by CEHM. This can be seen by noting that most of the posterior mass associated with the coefficient estimate for the "real" dummy is placed over negative values (fig.2).

[††] Geweke convergence diagnostic test is reported in the appendix.

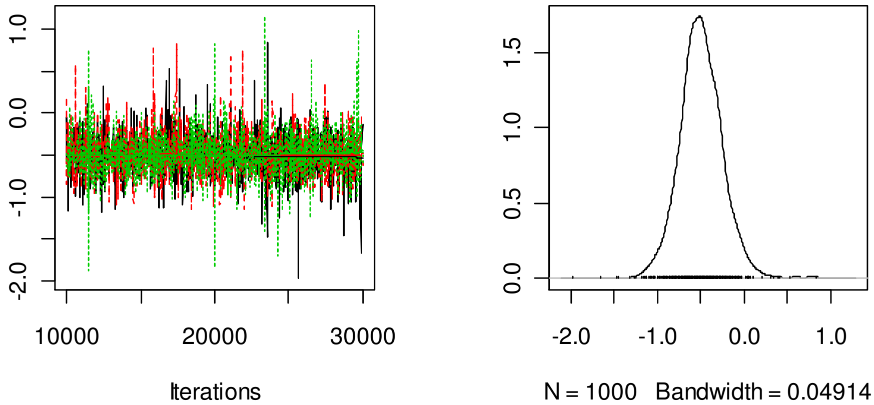

Figure 2. MCMC draws (3 chains) and posterior density for β_{real}

Nevertheless our results are also consistent with the intuition driving the analysis by HHW. That is, responses to hypothetical questions tend to have a larger error variance, as shown by the skeweness of the posterior distribution of the ratio σ_h/σ_r displaying a mean value of 1.42 and a median value of 1.17. However, the values of this distribution are not consistent with the very low scale factor ratio used by HHW to rescale the real payment subsample, which is likely to be a consequence of the lack of identification. The Bayesian analysis, we argue, provides a more reasonable estimate for the ratio between the two standard deviations, although the ratio is not so distant from one, since the 2.5 percentile is 0.43, and the 97.5 percentile is 4.09.

An overall measure of fit of the model is provided by the percentage of correctly estimated yes/no responses compared with those obtained by drawing from i.i.d. Bernoulli random variables with the θ parameter set to the sample average of yes responses in the data. The distribution of the hit rate for the probit model only partially overlaps with the one obtained from the Bernoulli draws as shown by figure 3. This seems to confute the statement by HHW that "the experiment reported by CEHM provide both real and hypothetical responses that for an overwhelming number of respondents are statistically indistinguishable from a coin flip".

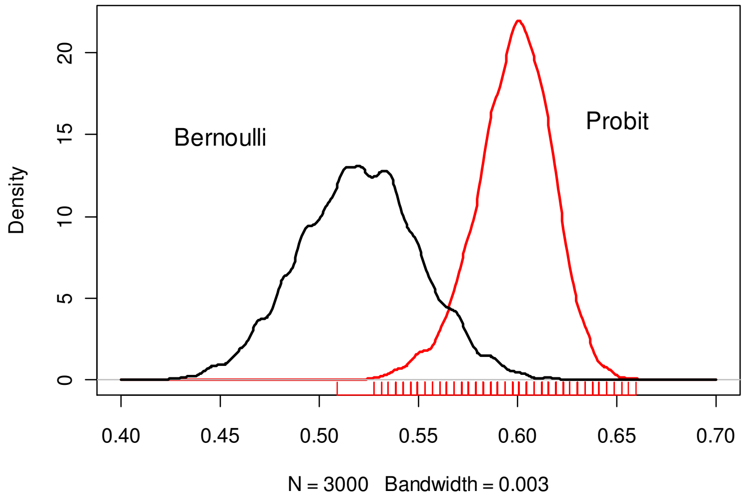

Figure 3 Distribution of model hit rate compared to a distribution of hit rate obtained drawing prediction for y from a Bernoulli distribution with parameter $\theta = \sum_{1}^{n} y_i / n$.

Finally, we explored the posterior distribution of the marginal effect of the real dummy on the probability to choose "yes", as suggested in Harrison (2006a, 2006b). The effect is calculated for each draw as the average over all individuals of the sample of the difference between the probability of obtaining a "yes" when the dummy takes the values 1 and 0, respectively:

$$\Delta \Pr = \sum_i \frac{1}{n} \left\{ \text{Prob}\left[y_i = 1 \mid \mathbf{x}_i, real = 1, \sigma = \sigma_r \right] - \text{Prob}\left[y_i = 1 \mid \mathbf{x}_i, real = 0, \sigma = \sigma_h \right] \right\} \quad (25)$$

Figure 4 illustrates that there is good evidence that the marginal effect of a real treatment on the probability of a "yes" vote is negative and different from zero, endorsing the findings by Harrison (2006a, 2006b).

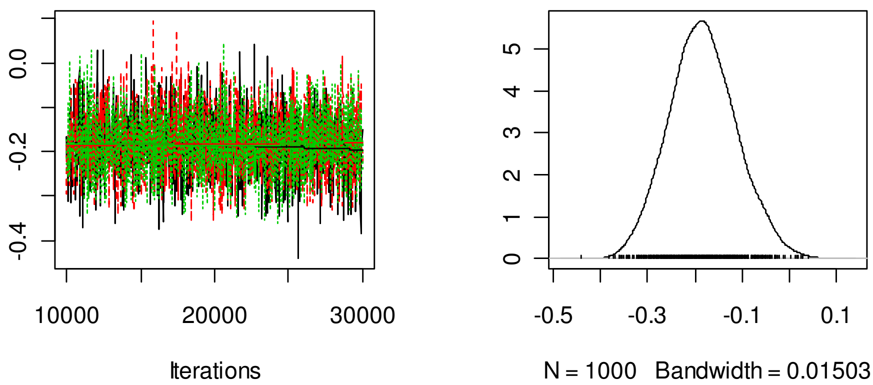

Figure 4. MCMC draws (3 chains) and posterior density for Marginal effect of treatment.

4. Conclusions

Since its publication the CEHM study has triggered an interesting debate over the reliability of the results initially reported by the authors. As far as the methodology is concerned, the initial debate was related to the issue of heteroskedasticity of the data conditional on the received treatment and the way to overcome the lack of identification of the relative scale parameters in single bid referendum data. The models employed so far (Carlsson and Johansson, 2010; Haab, Huang and Whitehead, 1999; Harrison, 2006a, 2006b) have provided mixed evidence of the existence of a significant impact of the treatment (real vs hypothetical) variable on the probability to answer "yes" to a CVM referendum question. Moreover, the poor conditioning of the dataset employed by CEHM hinders identification of both treatment and scale parameters (Carlsson and Johansson, 2010).

We proposed a Bayesian approach that overcomes the identification issue in these conditions, but at the cost of defensible assumptions on prior distributions of the scale parameters. Bayesian models are particularly suited to address problems of near-unidentifiability as the use of even relatively diffuse priors provides sufficient structure to carry out the estimation. In our case the Bayesian analysis seems to support the view by CEHM and Harrison (2006a, 2006b) that the ratio of "yes" to "no" responses is, to some extent, different across treatment conditions and that the data collected in the experiment by CEHM has some explicative power, at least if compared with a simple series of Bernoulli random draws.

5. References

J.H Albert, S. Chib (1993). Bayesian Analysis of Binary and Polychotomous Response Data, Journal of the American Statistical Association, 88 (422): 669-679.

T.A. Cameron, M.D. James (1987). Efficient Estimation Methods for “closed-ended” Contingent Valuation Surveys, Review of Economics and Statistics, May, LXIX(2): 269-276.

F. Carlsson, O. Johansson-Stenman (2010). Scale Factors and Hypothetical Referenda: A Clarifying Note, Journal of Environmental Economics and Management, 59: 286-92.

R.T. Carson, T. Groves (2007). Incentive and informational properties of preference questions, Environmental and Resource Economics, 37: 181-210.

R. G. Cummings, S. Elliott, G.W. Harrison, J. Murphy (1997). Are Hypothetical Referenda Incentive Compatible?, Journal of Political Economy, 105(3): 609-621.

R.G. Cummings, G.W. Harrison, E.E. Rutström (1995). Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?, American Economic Review, 85(1) March: 260-266.

A. Gelman, J.B. Carlin, H.S. Stern, Rubin (2004). D.B. Bayesian Data Analysis, Chapman and Hall, Boca Raton.

J. Geweke (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments, in: J.M. Bernado, J.O. Berger, A.P. Dawid, A.F.M. Smith (Eds.), Bayesian Statistics, Clarendon Press, Oxford, UK.

W. Greene (2002). Nlogit Version 3.0 Reference Guide, Econometric Software, New York.

W. Greene (2004). Econometric Analysis, Fourth Edition, Prentice-Hall, New Jersey.

T.C. Haab, J.C. Huang, J.C. Whitehead (1999), Are hypothetical referenda incentive compatible? A comment, Journal of Political Economy, 107(1): 186-196.

G. Harrison (2006a). Experimental Evidence on Alternative Environmental Valuation Methods, Environmental and Resource Economics, 34: 125–162.

G. Harrison (2006b). Making Choice Studies Incentive Compatible, in: B. J. Kanninen (Ed.) Valuing Environmental Amenities Using Stated Choice Studies, , Springer, Dordrecht: 67-110.

G. Harrison (2008), Personal communication.

E.E. Leamer (1978), Specification Searches, Wiley, London.

J.A. List, A.C. Gallet (2001), What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values?, Environmental and Resource Economics, 20(3): 241-254.

J.Little, Berrens, J.(2004). Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta−Analysis, Economics Bulletin, 3(6) : 1−13.

M. Plummer, N. Best, K. Cowles, K. Vines (2007). Coda: Output analysis and diagnostics for MCMC, R package version 0.13-1.

P.E. Rossi, G.M. Allenby, R. McCulloch (2005). Bayesian Statistics and Marketing, Wiley, Chichester.

V. K. Smith (1999). Of Birds and Books: More on Hypothetical Referenda, Journal of Political Economy, 107: 197–200.

J. Swait, J. Louviere (1993). The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models, Journal of Marketing Research, 30(3): 305-314.

A. Zellner (1971). An Introduction to Beyesian Inference in Econometrics, Wiley, New York.

6. Appendix

Tab. A1 Geweke Convergence Diagnostic Test

	Chain 1	**Chain 2**	**Chain 3**
ones	0.15	1.67	-1.89
real	-0.77	0.63	1.43
rh	0.20	2.23	1.62
age	-1.62	0.88	-0.54
male	-0.43	0.40	0.57
race	0.08	2.64	-0.77
income	0.31	-1.14	-0.21
married	1.27	-2.81	1.45
earn	0.16	-1.97	2.06
number	-1.25	0.21	-0.44
student	0.65	-2.02	0.15
s21	0.29	0.18	-0.02

Note: Fraction in first window =0.1, fraction in second window= 0.6