Journal of Environmental Economics and Management 59 (2010) 286–292

Contents lists available at ScienceDirect

Journal of Environmental Economics and Management

journal homepage: www.elsevier.com/locate/jeem

Scale factors and hypothetical referenda: A clarifying note

Fredrik Carlsson *, Olof Johansson-Stenman

Department of Economics, School of Business, Economics and Law, University of Gothenburg, Box 640, 405 30 Göteborg, Sweden

ARTICLE INFO

Article history:
Received 11 October 2008
Available online 7 March 2010

JEL classification:
C25
C91
C72

Keywords:
Hypothetical referenda
Hypothetical bias
Scale parameter
Contingent valuation

ABSTRACT

In this note we explore in detail the importance of, and problems associated with, correcting for variance differences between data sets obtained from hypothetical and real referenda. We show that a previous discussion in the literature rests on a problematic estimation of the relative scale factor. The implications are illustrated with data from Cummings et al. (1997) [5], as well as with simulated data. Moreover, we propose a concrete methodology for how to analyze cases where it is difficult, or even impossible, to estimate the relative scale factor due to informational limitations, such as when there is no variation of the bid. We conclude that it is valuable to be able to separate behavioral differences into variance differences and parameter differences in the underlying objective function. Yet, we argue that when using the results to interpret the results of other hypothetical referenda, it is sometimes the net effect, i.e., without correction for scale differences, that matters.

1. Introduction

Whether hypothetical referenda are valid in the sense of mimicking real referenda is important for policy evaluations. For example, contingent valuation studies of non-market goods are often formulated as a referendum. There is no sign of consensus about whether hypothetical referenda actually mimic real referenda. Some studies find a strong hypothetical bias, while others do not; see for example [5–7], and also see [13] for a meta-analysis.

The paper by Cummings et al. [5] (henceforth CEHM) is an important study that compared two treatments: one hypothetical and one real referendum directed towards people living close to a contaminated area. The scenario description in both treatments told the subjects that if everybody paid 10 USD, the amount of money would be sufficient to produce and distribute a "citizens' guide" that would provide valuable information about safe groundwater. In the hypothetical referendum, 45% voted yes and 55% voted no, whereas in the real referendum, 27% voted yes and 73% voted no. This sizable difference was found to be statistically significant, implying that they rejected the hypothesis that the hypothetical referendum is incentive compatible. However, in a comment, Haab et al. [8] (henceforth HHW) dispute this conclusion and claim that "the results of the experiments by Cummings et al. do not reject the hypothesis of incentive compatibility of hypothetical referenda" (p. 186). They further claim that if one corrects for a difference in variance between the two treatments, then there is no significant difference between them anymore.

In this note we explore in detail the importance of and problems with correcting for a possible difference in variance between data sets. The point raised by HHW is indeed potentially very important. However, as we will show, the way they identify and correct for the relative scale factor is inappropriate, and it may indeed be difficult or even impossible to identify a difference in variance when the informational basis for such an estimation is weak, such as in the case with the

* Corresponding author. Fax: +46 31 7861043.
E-mail address: fredrik.carlsson@economics.gu.se (F. Carlsson).

0095-0696/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jeem.2010.02.001

CEHM data. Independently of the present study, Stefani and Scarpa [15] also question the conclusions by HHW, albeit from a somewhat different perspective; they propose an alternative Bayesian approach to deal with the identification problem. Simply put, they achieve identification by making assumptions on the prior distributions of the relative scale parameter.

2. Referenda and between-sample heteroskedasticity: the CEHM data

Since discrete choice data provide limited information, identification problems can arise. In particular this concerns the identification of the variance of the latent variable, which becomes a central problem when one wants to compare or pool different data sets. Moreover, failure to correct for true underlying heteroskedasticity implies inconsistent parameter estimates, contrary to conventional continuous regression models; see for example [12,17]. In this section we first replicate the baseline CEHM results (based on their data). We implicitly assume that there is no between-sample variance difference between the hypothetical and the real treatments. Then we explain why HHW's attempt to first estimate the relative scale factor and then correct for this difference is inappropriate, and leads to biased estimates of possible real underlying sample differences. Finally, we outline how the estimation of the relative scale factor in principle should be conducted, and why this turns out to be more or less impossible due to informational limitations in the CEHM data.

2.1. The baseline CEHM results

There are two treatments in CEHM: one hypothetical (H) and one real (R). In both cases, participants are asked to vote yes or no to a particular program. CEHM estimate the probability of voting yes is estimated with a binary probit model:

$$P(\text{yes}) = F\left(\frac{\alpha+\beta^R+\beta x}{\sigma}\right), \tag{1}$$

where F is the cumulative density function for a standard normal distribution, α/σ is an intercept, β^R/σ is a corresponding shift parameter reflecting one type of difference between the treatments, x is a vector of socio-economic characteristics, and β/σ is the corresponding parameter vector. The shift parameter represents one type of behavioral difference between the two treatments. It could for example measure the extent of warm glow in the hypothetical referendum or the fact that respondents do not consider the cost to the same extent in the hypothetical referendum. Implicit in this formulation in (1) is that the probability of voting yes in the real and the hypothetical referenda, respectively, can be written as follows:

$$P^R(\text{yes}) = F\left(\frac{\alpha+\beta^R+\beta x^R}{\sigma^R}\right) \tag{2}$$

and

$$P^H(\text{yes}) = F\left(\frac{\alpha+\beta x^H}{\sigma^H}\right), \tag{3}$$

where it is assumed that the variance is the same, i.e., that *the relative scale factor equals one*, $\sigma=\sigma^R/\sigma^H=1$. The relative scale factor represents the other type of behavioral difference between the two samples. What CEHM assume is hence that there is no between-sample difference in variance. Our replication of CEHM's estimation results are presented in Table 1, column 1. We note that the parameter associated with the real referendum is highly significant.

2.2. Correcting for heteroskedasticity with the CEHM data

It is easy to see that there can be a potentially very large systematic effect of between-sample heteroskedasticity: if the variance is sufficiently large (theoretically approaching infinity), the responses are completely random. Then approximately 50% vote yes and 50% vote no, *irrespective of the respondents' true underlying preferences*. Consequently, it is conceivable that the fact that as many as 45% voted yes in the hypothetical case (compared with 27% in the real case) is due to a larger random component in the hypothetical case, perhaps since respondents did not bother to think as hard as they would have had it been an actual situation. If we could correct for the differences in variance, then we might find that there is no *remaining* significant difference.

In order to correct for differences in variance, $\sigma=\sigma^R/\sigma^H$ can be estimated either simultaneously or with a simple grid search procedure [4,16]; HHW used the grid search procedure suggested by Swait and Louviere [16]. However, HHW implicitly assume that there is no difference in the underlying function between the two data sets when estimating the relative scale parameter; in other words they estimate the relative scale parameter under the assumption that $\beta^R=0$. They then test, using a likelihood ratio test, whether the restricted pooled model can be rejected in favor of two separate models for the two samples. Performing this test they use the relative scale factor estimated as described above. They find that they cannot reject the restricted pooled model. In other words, they cannot reject the hypothesis that the differences between the samples are due to between-sample variance differences. This is a valid and important conclusion. We replicate this result in Table 1, column 2. As can be seen $\sigma \approx 0.04$, i.e., the estimated standard error is about 25 times larger

Table 1
Specifications of referendum preference and relative scale parameters; *t*-values in parentheses. In the simulated data the true underlying parameter values are as follows: constant = −0.5, real = −0.5, age = 0.01, sex = −0.1, scale = 1. Moreover, the true underlying yes-fractions are the same as the ones observed in the CEHM data, i.e., 27% and 45%, respectively.

	Replicated estimations based on the CEHM data			Estimations based on simulated data		
	CEHM specification	HHW specification		HHW specification		Simultaneous estimation
		Step 1	Step 2	Step 1	Step 2	
Constant	−0.6845 (−0.30)	−3.5477 (−0.68)	−4.7655 (−0.79)	−1.0707*** (−13.41)	−1.0104*** (−10.10)	−0.5102*** (−6.62)
Real	−0.4925** (−2.45)		1.2073 (0.39)		−0.0654 (−0.98)	−0.4959*** (−16.31)
Age	0.0129 (1.42)	0.0114 (0.71)	0.0114 (0.71)	0.0120*** (24.38)	0.0120*** (23.86)	0.0101*** (28.28)
Sex	−0.0906 (−0.53)	−0.2908 (−0.84)	−0.2891 (−0.84)	−0.1207*** (3.30)	−0.1210*** (−3.08)	10.1009*** (−3.79)
RH	0.1556 (0.62)	−2.7818 (−0.18)	−1.5849 (−0.34)			
Race	−0.0209 (−0.11)	0.3958 (0.82)	0.3951 (0.82)			
Income	−0.0026 (−0.53)	0.0062 (0.62)	0.0061 (0.61)			
Married	0.1251 (0.61)	−0.1506 (−0.36)	−0.1531 (0.37)			
Earn	0.0053 (0.03)	0.1853 (0.45)	0.1854 (0.45)			
Number	−0.0012 (−0.06)	−0.0115 (−0.23)	−0.0103 (−0.20)			
Student	0.1630 (0.55)	0.2892 (0.58)	0.2847 (0.57)			
Relative scale		0.0405 (0.23)		0.2557*** (9.18)		1.0071*** (4.41)
P (relative scale ≠ 1)					0.000	0.600
Log-*L*	−178.276	−178.986	−178.909	6243.99	6243.45	6232.25
N	286	286	286	10,000	10,000	10,000

***Statistically different from zero at the 1% significance level.
**Statistically different from zero at the 5% significance level.

in the hypothetical case, implying that the variance differ by a factor larger than 600. This appears implausible. Indeed, because the standard error for the real sample is non-negligible, choices in the hypothetical treatment would be close to pure noise. This is inconsistent with typical previous findings in environmental valuation studies [1,4].

Still, more problematic is that HHW then impose this estimated scale parameter in a second stage where a shift parameter is also estimated. These results are replicated in column 3, where it is found that β^R is not significantly different from zero. Following this, HHW claim, inappropriately in our view, that "when properly scaled, the real and hypothetical data are consistent with each other." The log-likelihood value is actually *lower* in the final estimated model than in the original CEHM model.[1] In order to see why this misspecification is crucial, let us return to the first stage by HHW. There are essentially no explanatory variables that can explain the differences between the samples, since all associated parameters are insignificant at the 15% level. Therefore we would expect the estimated relative scale factor here to be smaller than one in order to encompass the lower yes-response rate in the hypothetical case, irrespective of any differences in underlying variance. Indeed, in the extreme case where the explanatory variables do not explain anything at all, this "relative scale factor" can easily be identified by the yes-response fractions alone.[2] However, these fractions do of course not convey any information about variance differences.

Thus, we have argued that the β^R estimate will be highly biased by the above procedure. This mimics the warning by Swait and Louviere [16] (footnote 1): "Our procedure assumes that [...] systematic sources of differences across decision-makers have been accounted for by the inclusion of the appropriate variables in the utility functions. If this assumption is false, our procedure is inappropriate." It is then natural to ask what happens if we estimate the relative scale factor and β^R simultaneously, based on the original CEHM data? Since there is no variation of the "bid," i.e., everybody votes yes or no to a single value (10 USD), the only basis for calculating the relative scale factor is a comparison of how the yes-responses vary with respect to the explanatory variables. Thus, if there were no explanatory variables at all, it would be theoretically impossible to estimate the relative scale factor.[3] However, as mentioned, for the CEHM data all parameters associated with the explanatory variables are insignificant at the 15% level. This means that the informational basis for estimating the relative scale factor is very poor. Indeed, when performing a grid search it turns out that the likelihood function is

[1] Moreover, it is difficult to argue in favour of the HHW specification also at their step 1, since the CEHM model has a better fit and the same number of parameters.

[2] In the logit case, the probability of voting yes in the real referendum can without any explanatory variables be written as $1/(1+\exp(-\alpha^R))$, where the intercept α^R will be chosen so that the yes-probability will equal the yes-voting share of the respondents. In the final analysis of CEHM and HHW, this share is about 0.26. The corresponding expression for the hypothetical referendum is given by $1/(1+\exp(-\alpha^H))$, which will then equal about 0.46. We can then obtain the relative scale factor as $\sigma = \sigma^R/\sigma^H = \alpha^R/\alpha^H \approx [\ln((1-0.46)/0.46]/[\ln((1-0.28)/0.28] \approx 0.179$. However, CEHM as well as HHW estimate a probit (instead of a logit) for which there exists no algebraic solution. Still, it can be shown that the relative scale factor will typically be similar to the logit case [2]. In this case, $\sigma \approx 0.182$.

[3] Thus, it is possible, as we will also illustrate in the simulations, to identify both the shift parameter and the relative scale parameter, provided that the pattern with respect to other explanatory variables provides sufficient information. See for example [4], where data from an actual dichotomous choice referendum with no variation in the cost is pooled and compared with stated preference data. They are able to identify the relative scale parameters by imposing a particular functional form and using variation in socio-economic characteristics.

monotonically increasing in the relative scale factor (at least up to 100,000); i.e., the result implies the opposite of the HHW results, namely that the standard deviation is larger in the real referendum. However, the likelihood function is extremely flat, and by conducting a simple likelihood ratio test between $\sigma=10{,}000$ and 1, we cannot reject that the restricted model ($\sigma=1$) is the correct one, i.e., we cannot reject homoskedasticity so that $\sigma=\sigma^R/\sigma^H=1$. This should come as no surprise given the type and quality of the data, implying a poor informational basis for such an estimation.

3. A simple Monte-Carlo simulation

So far we have shown that it is potentially important to correct for between-sample heteroskedasticity, that the way HHW attempt to estimate the relative scale factor is inappropriate, and finally that it appears impossible to perform a correction for heteroskedasticity in an appropriate way with the CEHM data. In order to further illustrate the implications of correcting for scale differences in the appropriate and the inappropriate way, respectively, we now use an ideal simulated data set that mimics the real and the hypothetical referenda, where we thus know the true underlying parameter values.[4] Let the true underlying yes-fractions be the same as the ones observed in the total sample, i.e., 27% and 45%, respectively, but let the number of voters be 10,000 divided equally between the real and hypothetical referenda. This is achieved by assuming that the probability of voting yes in the real referendum is given by

$$P^R(\text{yes}) = 1 - F\left(\frac{\alpha+\beta^R+\beta_1\,\text{age}+\beta_2\,\text{sex}}{\sigma^R}\right), \tag{4}$$

where $\alpha=-0.5$, $\beta^R=-0.5$, $\beta_1=0.1$, $\beta_2=-0.1$, and $\sigma=\sigma^R/\sigma^H=1$. For expositional clarity we assume that there is no difference in variance between the two samples; the only difference is the shift variable.[5] While we set the relative scale parameter to unity, the method used by HHW is also inappropriate when the relative scale parameter is different from one.

Let us now use the methodology adopted by HHW on this simulated data.[6] We present the first step where we estimate the relative scale factor in Table 1, column 5. As shown, the relative scale factor is again very small, despite the fact that the true relative scale factor is unity. Moreover, when imposing this scale factor in the second step, reported in column 6, the coefficient of the dummy variable reflecting the real referendum is not significant. Finally, in column 7, we estimate the scale factor in the correct way, i.e., we simultaneously estimate the dummy variable for real referendum and the relative scale factor. Not surprisingly, the estimates here are very close to the correct values. Although we cannot draw the conclusion that the methodology of HHW will always lead to an insignificant referendum dummy parameter in the second step,[7] it will likely lead to a biased estimate of that parameter.

As we have indicated, one problem with the CEHM data is the limited information that the data contains, implying that it is difficult to know the extent at which the observed differences in yes-votes can be attributed to variance differences. To shed some further light on this issue consider the following question: Suppose that we know that β^R is zero, what is then the probability that we would observe a difference in yes-votes that is equally large or larger than the one observed by CEHM, i.e., 18 percentage points or larger? To answer this question we conduct a second Monte Carlo simulation, where we set β^R to zero and use the same sample sizes as in CEHM. The intercept is set to -0.65, which corresponds to a probability of yes-vote of 27%. We let the ratio $1/\sigma=\sigma^H/\sigma^R$ vary from 0.5 to 6, with steps of 0.2, where we make 10,000 draws for each level of the relative scale parameter.[8] For each draw we calculate the difference in the shares of yes-votes between the two real and hypothetical samples, and calculate the mean values. Fig. 1 presents the results of the simulations. We report both the average difference in the shares of yes-votes and the probability that the difference in shares is larger than 0.18, as a function of σ^H/σ^R.

As expected, when σ^H/σ^R is smaller than one (i.e., when the variance is larger for the real than for the hypothetical sample), the expected difference in yes-votes is negative, meaning that the expected share of yes-votes is larger in the real than in the hypothetical case. Naturally, when σ^H/σ^R increases, the expected difference in yes-votes and the probability that this difference exceeds 18 percentage points, increase. For example, when $\sigma^H/\sigma^R=3$, the average difference in yes-votes is almost 0.16, and the probability that the difference is larger than 18 percentage point is approximately 0.33. Similarly, when σ^H/σ^R is larger than 4.25, it is more likely than unlikely that we observe a measured difference in share of yes-votes of more than 18 percentage points. Overall, the illustration shows that the probability to observe an as large difference in shares as in CEHM is only high when there is a substantial between-sample differences in the underlying

[4] As noted by a referee, this simulation exercise can be seen as a thought experiment; see [11] for an interesting discussion of the relations between thought experiments, laboratory experiments and field experiments.

[5] For simplicity we include only two socio-economic variables: age and sex. These variables are independent; age has a normal distribution with mean 40 and standard deviation 10, whereas sex is a discrete variable that is either 0 or 1 with probability 0.5.

[6] Each model is estimated 100 times, and the averages are reported. The model is very simple, the data are well behaved, and the sample size is large, so there are very small differences between the individual simulations.

[7] If the other explanatory variables would explain a larger part of the variation, then the estimated scale factor would be less biased and correspondingly the estimated shift variable would be less biased in the second stage.

[8] Thus, we choose for presentational reasons to express the between-sample heteroskedasticity as the inverse of the relative scale factor as previously defined, i.e., as $1/\sigma=\sigma^H/\sigma^R$. The reason is that we are mainly interested in what happens when the variance of the hypothetical sample is larger than the variance of the sample involving real money, and it is more difficult to make graphical illustrations of ratios that go to zero.

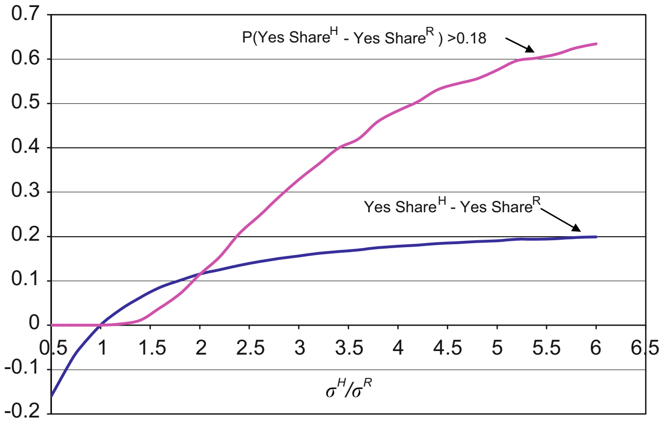

Fig. 1. The difference in expected shares of yes-votes, and the probability that the difference is larger than 0.18, as a function of σ^H/σ^R.

variances At the same time, the σ^H/σ^R ratio does not have to be very large in order to obtain the CEHM yes-vote difference with a non-negligible probability. For example, if $\sigma^R/\sigma^H=2$, we would obtain the CEHM yes-vote difference, or larger, with a probability of about 12%. Overall, if there is a considerable difference in the underlying variance between the two data sets, it is not unlikely to observe differences in behavior similar to the ones in CEHM.

4. When data does not allow for identification of the scale factor

The previous section demonstrated how we can (and how we should not) estimate the relative scale factor based on an almost ideal data set where we know the true underlying parameters. In principle it is possible to almost obtain such a data set using a laboratory experiments with a large sample size. In the CEHM case it would have been possible to run the experiment with both treatments using varying bid (cost) levels in the lab. Yet, while we encourage further experimental work on this issue, one must take into account the possibility that people's behavior in the lab may differ from the behavior outside the lab with the same monetary incentives, since a lab environment per se might sometimes affect people's behavior to a non-negligible extent [11]. In naturally occurring referenda there is, for obvious reasons, usually no variation in the "bid": people are simply asked to vote yes or no. Overall, in many, perhaps most, instances the data sets that are actually available are of much poorer quality than our simulated data. Here we will therefore briefly discuss how we might deal with situations such as with the CEHM data, where it is difficult to estimate the relative scale factor.

The fact that we cannot reject the homoskedastic model based on the CEHM data does of course not imply that it is correct or even a reasonable approximation. Indeed, there are plausible reasons for why one may expect the variance to be larger in the hypothetical case [14]. In order to deal with this problem, we propose the following methodology for this type of data, i.e., when we have a binary dependent variable (reflecting anything of interest) and no explanatory variables (or variables with poor explanatory power): *Calculate the degree of statistical significance for the variable of interest conditional on different values of the relative scale factor.* Then one can make a judgment about the robustness of the finding. If we apply this to the CEHM data we find that the real referendum parameter is significant at the 10% level for all values of the scale parameter ($\sigma=\sigma^R/\sigma^H$) above 0.57, and at the 5% level for all values above 0.71. Correspondingly, if the standard deviation is more than 75% larger for the hypothetical referendum, then the real referendum parameter is not significant at the 10% level. We leave it to the readers to judge whether these variance differences have reasonable order of magnitudes in the present case.

Although it is in principle always valuable to be able to separate behavioral differences into differences in variance and differences in parameters in the underlying objective function, it is sometimes the net effect that matters. Consider the following example: A hypothetical referendum (or a poll) is undertaken to measure people's responses to a certain proposal, and 40% vote yes. The task is to judge whether the yes-voting fraction in a real referendum would also be about 40%, or whether it would be considerably smaller. To the extent that the insights from CEHM are at all generalizable, the logical conclusion is that it would probably be considerably smaller. This is independent of whether or not the differences in the CEHM data are due to scale differences. Indeed, the stated purpose of CEHM is to test whether "a voter's behavior is independent of the use of a real or hypothetical referendum mechanism" [5, p. 611]. The answer to this question appears to be yes, regardless of whether this is due to scale effects or not, since a simple non-parametric test would reject the hypothesis of equal distributions of yes-votes between the two samples. This argument is also made by Harrison, although expressed in a slightly different way [9, endnote 16; 10, endnote 12]. He argues that if we care about the effect of the experimental treatment on the probability to vote yes then we should care about the variance term as well, since the marginal effect is a function of both the mean and variance of the estimated parameter. Caring about the marginal effect on the probability to vote yes, and not only the parameter estimate, is basically the same as saying that we care about the net effect.[9]

Consider again the result by CEHM that 45% vote yes in the hypothetical referendum and only 27% in the real one, and suppose we know that the true real-parameters is zero. Then this difference in responses is due to a difference in variance. If we would estimate this model in the correct way according to [16], the parameter associated with the real dummy variable would be close to zero and insignificant. Again, it is important to emphasize that this result does not imply there is no systematic behavioral difference between real and hypothetical voting. Indeed, we know that there is a huge difference of 18 percentage points! Thus, we cannot say that there is no systematic behavioral difference between the two referenda on the basis of the real dummy variable alone. This is because the larger the variance, the closer we get to 50% yes-responses, which is a systematic behavioral difference between the referenda. Thus, even if the CEHM data would have been large and well behaved, so that the relative scale factor could have been estimated, and even if the estimation would have revealed that the whole difference was indeed due to variance differences, as claimed by HHW, one could still not refute the CEHM conclusion that there is a sizable and systematic difference in the voting behavior between the samples.

As noted by a referee, another even simpler way of testing for differences between hypothetical and real treatments in cases such as the one considered by CEHM, where the covariates are either missing or not informative, is to simply rely on a χ^2 test. As was shown already by CEHM, such a test reveals a significant difference between the samples. One may also consider different straightforward non-parametric tests for such comparisons.

On the other hand, if the task instead were to interpret the results from a large CV study of the referendum type with varying bids, where WTP for an environmental improvement was estimated, then it would clearly be essential whether or not people's different behavior in hypothetical and real referenda is explained by scale differences. If so then the WTP estimate would in principle be unbiased since the scale factor in the CV study can be identified by using the so-called Cameron approach [3]. If not then the WTP estimate would be biased upwards.

5. Conclusions

In this note we made four conclusions regarding the role of relative scale factors when comparing real and hypothetical referenda: (1) Some previous discussion in the literature rests on an improper estimation of the relative scale factor, which leads to biased results. (2) Sometimes the informational basis for a proper estimation of the relative scale factor is very weak, such as when there is no variation of the bid, suggesting that the methodology outlined by Cameron [4], and Swait and Louviere [16] provides little guidance in such cases. (3) We proposed a concrete methodology for analyzing such cases. The basic idea is to calculate backwards the relative scale factors for which the shift parameter is significant (at different significance levels). (4) We finally conclude that it is in principle always valuable to be able to separate behavioral differences into variance differences and parameter differences in the underlying objective function. Yet, we argue that when using the results to interpret the results of other hypothetical referenda, it is sometimes the net effect, i.e., without correction for scale differences, that matters. This implies that some central conclusions of CEHM survive regardless of whether or not their observed voting differences are due to scale differences.

Acknowledgments

We want to thank the editors (Chuck Mason and Martin Smith), two anonymous referees, Vic Adamowicz, Martin Dufwenberg, Timothy Haab, Glenn Harrison, Ju-Chin Huang, Ricardo Scarpa, Måns Söderbom, Gianluca Stefani, Joffre Swait, and John Whitehead for very useful comments and discussions. The usual disclaimer applies. We are also grateful to Glenn Harrison for providing us with the original data. Financial support from Sida to the Environmental Economics Unit at University of Gothenburg is acknowledged.

References

[1] W. Adamowicz, J. Louviere, M. Williams, Combining revealed and stated preference methods for valuing environmental amenities, J. Environ. Econ. Manage. 26 (1994) 271–292.
[2] T. Amemiya, Advanced Econometrics, Harvard University Press, Harvard, 1985.
[3] T.A. Cameron, J. Environ. Econ. Manage. 15 (1988) 355–379.
[4] T.A. Cameron, G. Poe, R. Ethier, W. Schulze, Alternative non-market value-elicitation methods: are the underlying preferences the same?, J Environ. Econ. Manage. 44 (2002) 391–425.
[5] R. Cummings, S. Elliot, G. Harrison, J. Murphy, Are hypothetical referenda incentive compatible, J. Polit. Economy 105 (1997) 609–621.
[6] R. Cummings, G. Harrison, E. Rutström, Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive compatible, Amer. Econ. Rev. 85 (1995) 260–266.
[7] P. Frykbolm, Hypothetical question modes and real willingness to pay, J. Environ. Econ. Manage. 34 (1997) 275–287.
[8] T. Haab, J.-C. Huang, J. Whitehead, Are hypothetical referenda incentive compatible? A comment, J. Polit. Economy 107 (1999) 186–196.
[9] G. Harrison, Experimental evidence on alternative environmental valuation methods, Environ. Res. Econ. 34 (2006) 125–162.

[9] However, in our case we were unable to estimate the relative scale factor (if done in an appropriate way), due to the data limitations as mentioned. Consequently, we cannot estimate the marginal effect of the experimental treatment.

[10] G. Harrison, Making choice studies incentive compatible, in: B.J. Kanninen (Ed.), Valuing Environmental Amenities Using Stated Choice Studies, Springer, Dordrecht, 2006, pp. 67–110.
[11] G. Harrison, J. List, Field experiments, J. Econ. Lit. 42 (2004) 1009–1055.
[12] N. Kiefer, G. Skoog, Local asymptotic specification error analysis, Econometrica 52 (1984) 873–886.
[13] J. List, C. Gallet, What experimental protocol influence disparities between actual and hypothetical stated values, Environ. Res. Econ. 20 (2001) 241–254.
[14] J. Louviere, D. Hensher, J. Swait, Stated Choice Methods, Cambridge University Press, Cambridge, 2000.
[15] G. Stefani, R. Scarpa, Solving the flip of the coin/unidentification dilemma: a Bayesian perspective to the incentive compatibility of hypothetical referendum CVM, Working Paper, Department of Economics, Waikato Management School, 2008.
[16] J. Swait, J. Louviere, The role of the scale parameter in the estimation and comparison of multinomial logit models, J. Marketing Res. 30 (1993) 305–314.
[17] A. Yatchew, Z. Griliches, Specification error in probit models, Rev. Econ. Statist. 67 (1985) 134–139.