Is stated preference certainty individual-specific? - An empirical study

Jan-Erik Swärdh [a]

[a] *VTI - Swedish National Road and Transport Research Institute, Box 55685, SE-102 15 Stockholm, Sweden.*
E-mail: jan-erik.swardh@vti.se
The author is also affiliated to CTS - Centre for Transport Studies, Stockholm.

Abstract

The somewhat ad-hoc method of certainty calibration, based on self-stated preference certainty follow-up questions, has been found to be a successful method of eliminating or reducing hypothetical bias in stated preference studies. But is the preference certainty really context dependent, or do some subjects tend to always state themselves as certain regardless of the context, i.e. is the preference certainty dependent on a systematic unobservable individual-specific effect? This question is empirically analyzed in this paper using data where a preference certainty question follows a hypothetical willingness to pay question, in two different contexts. Estimated bivariate probit models provide no evidence for systematic individual-specific answers to the preference certainty follow-up questions of different contexts. Since there is no support for a randomly self-stated preference certainty either, this result is deemed to increase the credibility of certainty calibration.

JEL Classification C20 · C90 · D80

I. Introduction

It is argued that hypothetical bias is the main problem in the use of contingent valuation or other stated preference methods (Harrison, 2006). Furthermore, this bias seems to be relatively robust on elicitation method, context, and whether public or private goods are valued. Several meta studies show that

17 November 2010

the average hypothetical bias is severe (e.g. List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005).

To mitigate problems with hypothetical bias, different calibration approaches have been proposed. One such method is certainty calibration, where the information on individual self-stated preference certainty is taken into account in the empirical analysis. More specifically, the hypothetical choice is followed by a question where the subjects are asked to state their certainty of the previous hypothetical choice. Certainty calibration, despite its ad-hoc feature, has been shown to be a successful method for calibrating hypothetical values that are close to real values (e.g. Blumenschein et al., 2008; Champ et al., 1997; Johannesson et al., 1998).

Nevertheless, the lack of theoretical models for preference certainty raises the question on whether certainty calibration really fills the gap between stated choices and underlying preferences. Also, Akter et al. (2008) points out that the fundamental issue for further development of preference certainty, is whether this can be measured accurately. A problem being that observable individual characteristics, such as age, may influence preference certainty. For example, Svensson (2006) finds a weak tendency towards age having a positive effect on preference certainty. Moreover, the problem may become even more severe if there is an unobserved individual-specific effect that systematically influences preference certainty, which would imply that some individuals state a high certainty, independent of the given hypothetical choice. If some individuals were to repeatedly state that they were certain, regardless of the given context, this would capture a type of attitude-effect, and using preference certainty to close the gap between a single stated choice and underlying preferences would, in this case, break down. On the other hand, some individuals may be better at understanding the specific context and therefore better at answering hypothetical questions. This is exactly the finding in Champ and Bishop (2001) where attitudes and donation experience are found to influence the preference certainty. However, if the latter is the case, the inevitable objection is that input values in cost benefit analysis, based on stated preferences, would not be representative of the population, but would represent only individuals who are good at understanding and answering hypothetical questions.

This study is, to the knowledge of the author, the first to empirically analyze the relationship between two different preference certainty follow-up questions in economic valuation contexts, where both follow-up questions are addressed to the same subjects. The analysis is based on bivariate probit models where no correlation between the error terms of the two probit equations, is interpreted as a good sign for the credibility of certainty calibration. Also, the preference certainty should not comprise random answers, which can be empirically checked in the bivariate probit model, by including a variable that measures the absolute deviation between the estimated individual-specific willingness to pay (WTP) and the offer price.

Data from 85 students who participated in a lab experiment is used. The students answered a number of questionnaires during a one hour period and were compensated with 100 Swedish Crowns (SEK [1]) for this work. [2]

Within these questionnaires, two different hypothetical WTP questions fol-

[1] 100 SEK is approximately 10.7 EUR.
[2] More about this experiment can be found in the fourth Essay of Swärdh (2009).

lowed by a preference certainty question were posed. Firstly they were offered an option to leave the experiment 45 minutes early on payment of a predetermined amount of money, followed by a question of preference certainty with the alternatives ‘yes, completely certain’ and ‘no, not completely certain’. Secondly they were offered the option to pay for an item of traffic safety equipment that would eliminate fatal and severe injuries in Swedish traffic, followed by a ten-point-scale preference certainty question where 1 is ‘very uncertain’ and 10 is ‘very certain’.[3]

A nice characteristic of this experiment is that between the two preference certainty questions, was a time lapse of about 25 minutes, including a different questionnaire. The implication of which should be to lower the risk of any anchoring effects from the earlier preference certainty question on the answer to the latter preference certainty question. The different types of follow-up questions used may also help to avoid anchoring effects.

II. Results

In Table 1, descriptive statistics of the estimation sample are presented. Since the sample consists of students, both the average age and the average income is low, and age also has a low variation. In fact, most subjects are between 20 and 24 years of age. Furthermore, the average scaled preference certainty is

[3] Note that both yes-responders and no-responders are asked about the preference certainty, whereas common practice (e.g. Blumenschein et al., 2008) is to ask only yes-responders about preference certainty. By recoding or excluding only yes-responders, the estimated WTP with certainty calibration will necessarily be lower than the estimated WTP without certainty calibration. Therefore, asking both yes-responders and no-responders may be a more appropriate method. However, keep in mind that the objective of this study is not to estimate WTP by applying certainty calibration.

7.65 with a standard deviation of 2.07. This average coincides with the result of previous studies, which suggests that the restrictive sample of students in this study does not affect the distribution of the stated certainty in any systematic way.[4] In addition, the traffic risk reduction offer is accepted by about 73% of the subjects while the offer to leave the experiment early is accepted by only 13% of the subjects. About 71% of the subjects state that they are completely certain of their choice in accepting or rejecting the offer to leave the experiment.

[Table 1 about here]

In Figs 1 through 3, the distribution of the scaled preference certainty is presented for the complete sample, yes-responders and no-responders, respectively. Ten is the most commonly chosen certainty level by the complete sample and the yes-responders, whereas nine is most commonly chosen by the no-responders. Also, all distributions are skewed to the right, especially the distribution of yes-responders.

[Fig. 1 about here]

[Fig. 2 about here]

[4] In ten different samples based on four previous studies (Blomquist et al., 2009; Hultkrantz and Shengcong, 2009; Svensson, 2006; Vossler and McKee, 2006), the mean of the scaled preference certainty varies between 7.15 and 8.84.

[Fig. 3 about here]

In the empirical analysis, the scaled preference certainty is treated in accordance with the literature, that is setting a threshold value for when the subjects are treated as certain. This threshold value differs across studies and has usually varied between ten (e.g. Champ et al., 1997) and eight (e.g. Champ and Bishop, 2001). In a recent study, Blomquist et al. (2009) compare which preference certainty, on a ten-point scale, corresponds to a definitely sure response, and their finding is that equivalence is reached near ten on the certainty scale. Furthermore, a threshold value of nine will be tested in this study. Thus, all three of these threshold values are used to check the robustness of the results.

The first method used to test whether there is any association across the two different preference certainty answers, is the non-parametric gamma test by Goodman and Kruskal (1979). Table 2 shows the results of these tests where the qualitative preference certainty is compared to each of the codings of the scaled preference certainty, as described in the last paragraph. The results show that none of these tests are significant, which means that we cannot reject the hypothesis that there is no association between the two preference certainty responses.

[Table 2 about here]

Nevertheless, the relation between the two preference certainty responses may be influenced by controlling for other effects. Therefore, bivariate probit models (see e.g. Cameron and Trivedi, 2005, p. 522) are estimated. A bivariate probit model consists of two probit models where the error terms are assumed to be joint normal distributed with means zero, variances one and correlation ρ. In this study, a non-zero ρ is considered an indication of an individual-specific effect of preference certainty.

The procedure before the bivariate probit models can be estimated is as follows. First, WTP models are estimated according to the logit specification

$$p(\text{accept offer}|\pi, \mathbf{x}) = \frac{\exp(\gamma\pi + \mathbf{x}'\beta)}{1 + \exp(\gamma\pi + \mathbf{x}'\beta)} \tag{1}$$

where π is the offer price and $\mathbf{x}$ is a vector consisting of the socioeconomic variables *age*, *gender*, *married/cohabitant*, *education* defined as the *number of completed semesters of university education*, and *income*.

Subsequently, individual-specific WTP is estimated according to the formula (see e.g. Blumenschein et al., 2008, p. 126)

$$WTP_i = (-1/\widehat{\gamma}) \times \ln(1 + \exp(\mathbf{x}_i'\widehat{\beta})) \tag{2}$$

where $\widehat{\gamma}$ and $\widehat{\beta}$ are estimated parameters from Equation 1.

Finally, the absolute deviation between WTP and offer price is calculated as

$$DEV_i = |WTP_i - \pi_i| \tag{3}$$

which is included as a covariate in the bivariate probit models. The respondents are expected to be more uncertain about their choice the narrower the gap between their individual WTP and their offer price (following argumentation in Wang, 1997, for example). Thus, DEV is expected to be positively correlated with preference certainty.

In Table 3, the estimated results of the bivariate probit models are presented. Three different models are presented where the scaled preference certainty is treated differently. The absolute deviation has the expected positive influence for the traffic safety question and is also significant, at least at the 5%-level. For the value of time question, on the other hand, there is an unexpected negative parameter estimate for the absolute deviation. However, this effect is not significant at conventional significance levels.

Most interesting, though, is the correlation between the error terms of the probit models, which is estimated to be between -0.23 and -0.05 in these three models. Negative correlation implies that a certain answer to one of the WTP questions is likely to be followed by an uncertain answer to the other WTP question. This is not the expected relation, given an individual-specific effect, however, the p-values suggest far from significance at conventional significance levels. To sum up the results, the hypothesis of non-correlation between the answers to the two different preference certainty questions, cannot be rejected.

[Table 3 about here]

III. Concluding discussion

In this study, the objective was to analyze whether a systematic unobserved individual-specific effect, influences the self-stated preference certainty of a follow-up question to a hypothetical choice. From estimated bivariate probit models, no empirical support for such a systematic effect is found, although the small and restrictive sample used, implies that the results have to be interpreted with caution. Furthermore there is no support for complete random answers to these preference certainty questions either, and thus the conclusion is that the result of this study provides some support for the credibility of certainty calibration.

Nevertheless, certainty calibration may be criticized for other reasons, in particular the ten-point scale where there is no given threshold value choice for the decision where hypothetical subjects with less certainty should be recoded, or excluded from the empirical analysis. Therefore, it is difficult to choose the threshold value when certainty calibration is applied in a WTP-study, based on stated preferences. It appears, however, that in this study there is no difference in the relationship of the two different preference certainty questions, regardless of the coding used for the scaled preference certainty question.

As further research, more empirical studies in this field are required to support (or not support) certainty calibration. A larger sample, which is more representative of the population would be fruitful. Such studies may also use to their advantage the practice of not presenting the follow-up questions of preference certainty too close to each other, since this procedure is most likely good in reducing potential problems with anchoring effects. Finally, other effects that would be interesting to study, are learning effects, certainty calibration in repeated stated choice, and the consistency of the answers to preference certainty follow-up questions.

References

Akter, S., Bennett, J. and Akhter, S. (2008) Preference uncertainty in contingent valuation, *Ecological Economics*, **67**, 345–351.

Blomquist, G. C., Blumenschein, K. and Johannesson, M. (2009) Eliciting willingness to pay without bias using follow-up certainty statements: Comparisons between probably/definitely and a 10-point certainty scale, *Environmental and Resource Economics*, **43**, 473–502.

Blumenschein, K., Blomquist, G. C., Johannesson, M., Horn, N. and Freeman, P. (2008) Eliciting willingness to pay without bias: Evidence from a field experiment, *Economic Journal*, **118**, 114–137.

Cameron, A. C. and Trivedi, P. K. (2005) *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.

Champ, P. A. and Bishop, R. C. (2001) Donation payment mechanism and contingent valuation: An empirical study of hypothetical bias, *Environmental and Resource Economics*, **19**, 383–402.

Champ, P. A., Bishop, R. C., Brown, T. C. and McCollum, D. W. (1997) Using donation mechanisms to value nonuse benefits from public goods, *Journal of Environmental Economics and Management*, **33**, 151–162.

Goodman, L. A. and Kruskal, W. H. (1979) *Measures of Association for Cross Classifications*, Springer-Verlag, New York.

Harrison, G. W. (2006) Experimental evidence on alternative environmental valuation methods, *Environmental and Resource Economics*, **34**, 125–162.

Hultkrantz, L. and Shengcong, X. (2009) Mitigating hypothetical bias in Value of time Studies: Lab-experiment results, Discussion Paper, Swedish Business School, Örebro University, Örebro.

Johannesson, M., Liljas, B. and Johansson, P. (1998) An experimental comparison of dichotomous choice contingent valuation questions and real purchase decisions, *Applied Economics*, **30**, 643–647.

List, J. A. and Gallet, C. A. (2001) What experimental protocol influence disparities between actual and hypothetical stated values?, *Environmental and Resource Economics*, **20**, 241–254.

Little, J. and Berrens, R. (2004) Explaining disparities between actual and hypothetical stated values: Further investigation using meta-analysis, *Economic Bulletin*, **3**, 1–13.

Murphy, J. J., Allen, P. G., Stevens, T. H. and Weatherhead, D. (2005) A meta-analysis of hypothetical bias in stated preference valuation, *Environmental and Resource Economics*, **30**, 313–325.

Svensson, M. (2006) Estimates of the value of a statistical life from two Swedish surveys using the "certainty approach" calibration, Working Paper 2006:6, Department of Business, Economics, Statistics and Informatics, Örebro University, Örebro.

Swärdh, J-E. (2009) *Commuting Time Choice and the Value of Travel Time*, Ph.D Thesis, Örebro University, Örebro.

Vossler, C. A. and McKee, M. (2006) Induced-value tests of contingent valuation elicitation mechanisms, *Environmental and Resource Economics*, **35**, 137–168.

Wang, H. (1997) Treatment of 'Don't-Know' responses in contingent valuation surveys: A random valuation model, *Journal of Environmental Economics and Management*, **32**, 219–232.

Fig. 1. Distribution of the preference certainty scale responses

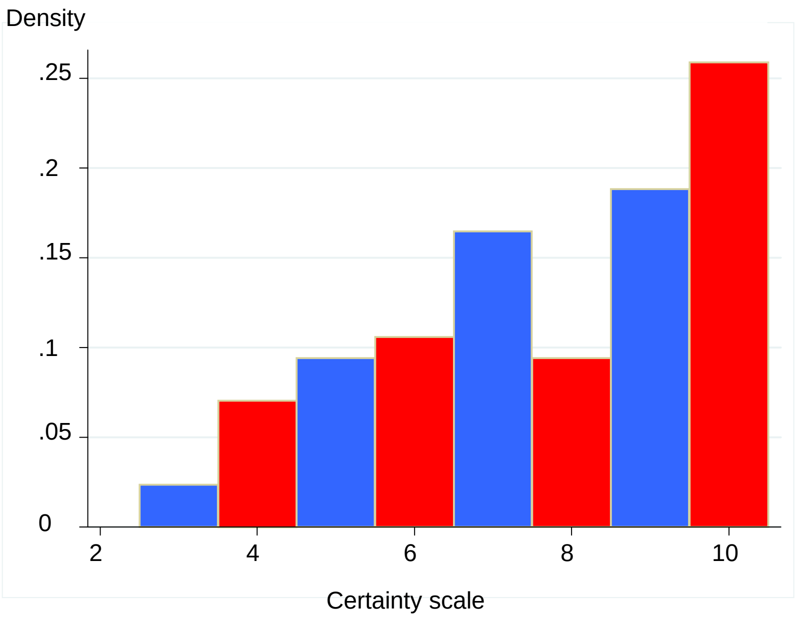

Fig. 2. Distribution of the preference certainty scale responses for yes-responders

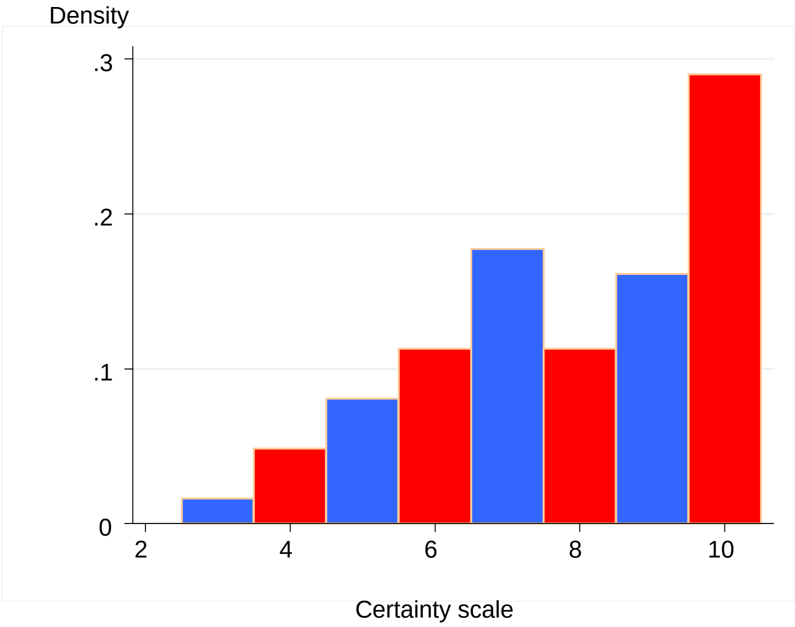

Fig. 3. Distribution of the preference certainty scale responses for no-responders

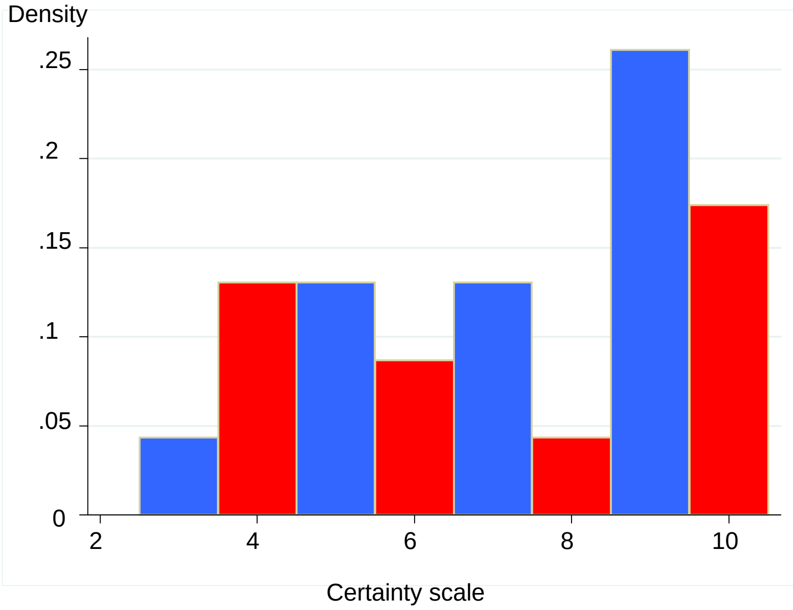

Table 1
Descriptive statistics

Variable	**Mean**	**SD**
Age	21.56	2.652
Female	0.424	
Married or cohabitant	0.153	
Income (SEK/month)	6413	3531
No university semesters completed	0.176	
University - 1 or 2 semesters completed	0.412	
University - 3 or 4 semesters completed	0.341	
University - at least 5 semesters completed	0.071	
Accepting traffic risk reduction offer	0.729	
Scaled preference certainty	7.647	2.074
Accepting offer to leave the experiment	0.129	
Qualitative preference certainty	0.706	
No. of observations	85	

Notes: SD of indicator variables is not shown since it is determined by the mean according to $\sqrt{\mu(1-\mu)}$, where μ is the mean.

Table 2
Tests of association between the two preference certainty responses using Goodman and Kruskal's gamma

Coding of Scaled Certainty	**Gamma**	**Asymptotic SE**
Certain $= 10$	-0.077	0.266
Certain ≥ 9	-0.094	0.237
Certain ≥ 8	0.172	0.234
No. of observations	85	

Table 3
Bivariate probit estimation

Variable	**Coefficient**	**SE**
Certain = 10		
DEV Traffic Safety	0.508*	0.245
DEV Value of Time	-0.150	0.094
Correlation	-0.225	0.232
Certain ≥ 9		
DEV Traffic Safety	0.479*	0.216
DEV Value of Time	-0.141	0.093
Correlation	-0.215	0.200
Certain ≥ 8		
DEV Traffic Safety	0.680**	0.224
DEV Value of Time	-0.147	0.093
Correlation	-0.054	0.197

Notes: **, * and † denote significance at the 1% level, 5% level and 10% level, respectively. Standard errors are robust. The number of observations is 85. *DEV* for traffic safety is multiplied by 1000, and *DEV* for value of time is multiplied by 10.