An addendum to: A Meta-Analysis of Hypothethical Bias in Stated Preference Valuation[1]

Stefani S.[1], Scarpa R.[2] and Lombardi G.V.[1]
1 Department of Economics and Management, University of Florence, Florence, Italy
2 Economics Department, Waikato Management School, Hamilton, New Zealand,
gianluca.stefani@unifi.it

Abstract

A recent study published by Murphy et al. (2005) reported results of a meta-analysis of hypothetical bias using 28 valuation studies. The authors found a median ratio of hypothetical to actual values of 1.35 but they did not investigate the ratio of scales of the hypothetical and actual value distributions, which is of great relevance in joint stated and revealed preference analysis. We propose an addendum to Murphy et al. (2005) to provide some insights on the distribution of the scale factor across 23 studies for which relevant data is available. We also describe a method to supply priors to future studies that use Bayesian approaches to model merged revealed and stated preference data.

Keywords: Contingent Valuation, experiments, scale identification, meta-analysis, stated preferences

JEL classification: C9, H41,Q26,Q28

Abbreviations: CF = calibration factor ; CV= contingent valuation; ICF= inverse of calibration factor; SF = scale factor ; WTA = willingness to accept; WTP = willingness to pay

1. Introduction

Stated preference methods are widely used in nonmarket valuation of environmental goods. However, they have been criticised for a number of reasons revolving around the issues of credibility and reliability of hypothetical responses (Cummings et al.

1997, Diamond & Hausman 1993, Green et al. 1998, to name but a few in the context of contingent valuation). The difference between responses in hypothetical and real settings, known as hypothetical bias, is an issue that has given rise to a fierce debate among scholars and has motivated much research effort. A key question in the ensuing research agenda has been the estimation of a calibration factor (CF). This is the ratio between hypothetical (stated preference) and actual (or revealed preference) values. Using CF values elicited with hypothetical choices may be corrected to obtain value estimates similar to those obtainable from revealed preference studies.

A methodological approach that has generated much attention in this area of applied research has been the use of meta-analysis. At least three literature review or meta-analysis studies have investigated the scope and extension of CF. Harrison and Rutstrom (2008) using 35 observations report a CF ranging from 0.75 to 26. List and Gallet (2001) analyse 29 studies with a total of 174 observations of willingness to pay (WTP) and willingness to accept (WTA) estimates. According to their study hypothetical values are about three times larger than real ones, with CF being larger for WTA rather than WTP or when the values are elicited for public rather than private goods. Little and and Berrens (2004) expanded the dataset of List and Gallet to include 17additional observation. The CF from their dataset ranges from 2.93 to 3.34 with a mediam value of 3.13. Murphy et al. (2005), drawing on the study by List and Gallet, proposed a new meta-analysis focussing on WTP estimates and including only observations that employ the same mechanism to elicit hypothetical and real values. The authors selected 28 studies that yield a total of 83 observations for which the distribution of CF is skewed with a mean value of 2.60 and a median value of 1.35. The authors found mixed results about the determinants of CF. Students and group setting seem to widen CF, while discrete choice format and valuation of private goods would have the opposite effect. In a cautionary note, the authors warn that results are sensible to model specification and that the choice of explanatory variables is affected by the lack of a theory explaining hypothetical bias.

All four studies purport the ratio between hypothetical and actual values as a key factor in criterion validity of stated preference estimates under the assumption that values elicited from revealed preference data are closer to the truth.

Our point of departure is the observation that distributions of ratios of value estimates are not completely characterised by location parameters alone (such as mean or median).

Dispersion parameters are also of crucial importance, especially in the context of joint preference estimation from merged revealed and stated preference data (e.g. Hensher, Louviere and Swait, 1999). In this context there are good theoretical reasons for the existence of a difference in error scale from different data sources, which has been corroborated by much empirical evidence (Louviere, 2001). Similarly Cameron et al. (2002) states that "What would be most valuable for predicting actual demand behavior from stated preference choice data would be some means of using common underlying systematic preference parameters, but mapping the dispersion parameter from the particular stated preference method into the likely corresponding dispersion parameter for a revealed preference choice context. This might allow prediction of the distribution of WTP for real market choices."

For example, the results from one of the first papers addressing the impact of real vs. hypothetical treatments on values elicited with contingent valuation (CV) of public goods (Cummings et al., 1997) were indeed questioned with respect to the assumption of equal variance across treatments two years later (Haab et al. 1999)

We define as inverse relative scale factor (IRSF) the ratio of standard deviations of real over hypothetical value distributions:

$$IRSF = \sqrt{\frac{\sigma_r^2}{\sigma_h^2}} = \frac{\sigma_r}{\sigma_h} \qquad (1)$$

where σ_r and σ_h are standard deviations with subscripts referring to "real" and "hypothetical" distributions respectively. We named the ratio inverse relative scale factor since the scale factor is usually defined as $\mu = 1 / \sigma$ (Adamowicz , Louviere and Williams, 1994) and the relative scale factor as $SF = \mu_r / \mu_h$ whilst our index is given by $IRSF = \mu_h / \mu_r$.

The aim of this note is hence to provide some specific insights on the distribution of the IRSSF between hypothetical and real value distributions from a subset of 23 studies out of the original 28 considered by Murphy et al.(2005), for which relevant data on scale is available. Our focus is on deriving estimates of the IRSF rather than exploring the determinants of hypothetical bias, therefore the note should be considered as an addendum rather than a comment to the original paper by Murphy.

The remainder of this note is set out as follows. Section 2 illustrates data, estimation procedures and results. Section 3 proposes a prior for error precision useful to analyse real and hypothetical WTP data. Section 4 provides a summary of findings and conclusions. We provide an assessment of the empirical distribution of the SF across a sample of stated preference studies finding that differences of variances are mild, a result similar to that provided for CF by Murphy et al. (2005). We found that CF and SF are correlated and factors that affect the former also tend to affect the latter. We also provide a possible parametric specification for the IRSF distribution.

1. Alternative measure of dispersion of WTP

As different measure of dispersion are available across the reviewed studies depending on the estimation framework adopted we provide below a simple model that helps clarifying the differences among alternative measures.

Let us start with a simple linear random WTP model:

$$WTP_i = x_i'\beta + \varepsilon_i \qquad (2)$$

A first important distinction is between the marginal or unconditional variance of WTP and the variance of the error term of the model:

$$VAR(WTP) = E_x[VAR(WTP|x)] + VAR_x[E(WTP|x)] \qquad (3)$$

or

$$VAR(WTP) = VAR(\varepsilon) + VAR_x(x\beta) \qquad (4)$$

Equation 4 decomposes the unconditional variance of WTP into the variance of the error term and the variance of the conditional mean of WTP with respect to x. It is clear that the ratio of unconditional variances will be always different from the ratio of the error term variances unless the x are fixed in the hypothetical and real treatment or the ratio of the $VAR_x(x\beta)$ is the same of the ratio of the $VAR(\varepsilon)$:

$$\frac{\sigma^2_{wtpH}}{\sigma^2_{wtpR}} \neq \frac{\sigma^2_{\varepsilon H}}{\sigma^2_{\varepsilon R}} \qquad (5)$$

Further measures of dispersion can arise as some researchers calculate fitted WTP for a representative person. Then drawing from the asymptotically joint normal distribution of the maximum likelihood parameter estimates they build up a sampling distribution of fitted WTP estimates following the procedure set out by Krinsky and Robb (1986) to estimate confidence intervals for elasticities. Then the distribution reflects the estimation precision for all of the parameters in the model not just the error precision, and shows how estimation efficiency affects the range of plausible values for WTP for a representative subject. In the case of the linear model 2 estimated through OLS we it is well know that the asymptotic estimator of VAR(b|X) is given by:

$$\widehat{VAR}(b|X) = \sigma_u^2 (X'X)^{-1} \qquad (6)$$

Therefore the variance of estimated WTP for a representative subject (at the mean values of x, $\bar{x}$) is:

$$VAR(b\bar{x}|X) = \bar{x}'[\sigma_u^2 (X'X)^{-1}]\bar{x} \qquad (7)$$

Which again is different from either VAR(ε) and its estimator VAR(u) or from (4).

2. Data, estimation and results

We supplemented the dataset employed by Murphy et al. (2005)[2] by recording measures of dispersion of the WTP distribution whenever available. We were able to collect data from 23 out of the original 28 studies providing 67 observations[3]. In addition we retrieve 4 more observation from two studies surveyed by Little and Berrens (2004). Overall, our dataset includes 25 studies and 71 observations.

In the augmented dataset available measures of dispersion can be classified according to both the type of measure of dispersions outlined in the previous section and the format the dispersion is provided with. We classified different formats for dispersion measure into 4 groups as follow.

1) *Standard deviation*. Studies based on experimental auctions and open-ended elicitation formats generally provide data on standard deviations of WTP values distributions.

2) *Confidence interval*. Most dichotomous choice and some open ended CV methods provide confidence intervals for the estimates of the mean WTP. As the sample size is similar for real and hypothetical treatments, the width of confidence intervals is proportional to the standard deviation of WTP. Therefore we maintain that the ratio of the sizes of confidence intervals is a close proxy for IRSF.
3) *Sigma*. Studies that employ dichotomous choice data often use probit or logit models to explain outcome probabilities. In such cases it is possible to recover the standard deviation σ of the underlying distribution from the inverse of the estimate of the parameter of the bid variable[4] as described in Cameron and James (1987). Actually, in the case of logit models the inverse of the parameter of the bid variable gives $\kappa = \sigma\sqrt{3}/\pi$, that is the dispersion of the error term in the logistic regression. However, being k a linear function of s the ratios are equal:

$$\frac{\kappa_r}{\kappa_h} = \frac{\sigma_r}{\sigma_h} \qquad (8)$$

4) Scale factor. Finally, a single study (Carlsson and Martinsson ,2001) carried out using multiple choices, reports directly the corresponding scale factor.

From each of the studies employed here a IRSF value is obtained by dividing the measure of dispersion of the real or actual subsample by the one estimated from the hypothetical subsample.

Table 1 Observations classified according to format and type of dispersion

	Type of distribution			
Format	Error	Parameters	Marginal	*Total*
Confidence interval	1	5	8	*14*
Scale_fact	1	0	0	*1*
Sigma	14	0	0	*20*
Stand. Dev	0	8	34	*36*
Total	*16*	*13*	*42*	*71*

Most of the observations in the dataset are estimates of marginal WTP distributions, formatted either as standard deviation or confidence interval. Only 16 observations provide an estimate of error distribution followed by the group of observations where the measure provided is a function of model parameters standard errors through Krinsky-Robb type procedures.

3. Results: summary statistics

Overall, the mean value of the IRSF in our sample is 0.67 with a standard deviation of 0.41. However, the IRSF distribution is quite different across the three distribution types confirming their different nature. All means and medians are smaller than 1, consistently with theoretical expectations. Haab et al. (1999) state that real experiments control more effectively sources of variability, therefore the distribution of elicited values is likely to be less dispersed than in hypothetical settings.

Table 2 Summary statistics of IRSF across distribution types

	Type of distribution			
Form	error	parameters	marginal	*Total*
min	0.45	0.04	0.07	*0.04*
max	1.85	1.40	1.51	*1.85*
mean	0.90	0.67	0.58	*0.67*
med	0.82	0.69	0.54	*0.66*
st.dev	0.39	0.50	0.35	0.41

It is worth noticing that with a similar number of observations, measures of IRSF based on all model parameters distributions are more dispersed than those based on model error distribution. Density estimates of the three distributions of IRSF are reported in figure 1[5].

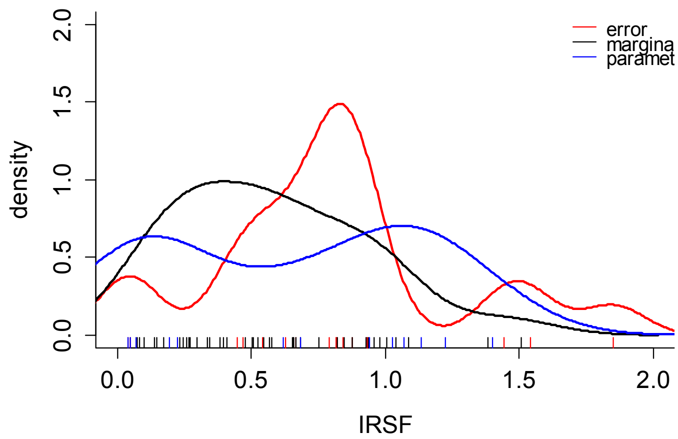

Figure 1 Density estimates of IRSF

IRSF shows an overall weak and positive correlation (r=0.58) with the inverse of CF (ICF), which is the ratio of actual vs hypothetical mean. It is the IRSF obtained from marginal distributions of WTP that shows the highest correlation with ICF. Interestingly, IRSF form error distributions does not seem to be correlated with ICF. However this results is likely to be affected by the presence of outliers as it can be seen from figure 2.

Table 3 Correlation between ICF and IRSF

	ρ	CI
error	-0.13	(-0.56 , 0.36)
parameters	0.66	(0.18 , 0.89)
marginal	0.81	(0.67 , 0.90)
all	0.58	(0.41 , 0.72)

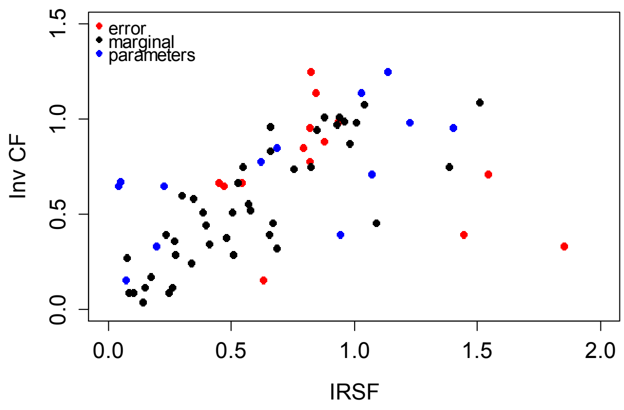

Figure 2 ICF vs IRSF

4. Results: regression analysis

To try and explain hypothetical bias we also regress IRSF on ICF and on those explanatory variables used by Murphy et al. (2005) (tab. 1). This allows us to see if there is any further marginal effect besides ICF showing significance.

Table 4. Regression results: marginal distribution type only

Estimate	Estimate	Std. Error	t value
(Intercept)	0.15	0.06	2.33
ICF	0.81	0.09	8.61
Choice	-0.08	0.09	-0.82
Private	-0.12	0.09	-1.28
Student	0.02	0.06	0.28
Within	0.19	0.09	1.95
Calibrate	0.19	0.08	2.46

Multiple R-squared: 0.77,

Adjusted R-squared: 0.73

We analyse only the largest group of observations derived from marginal distributions of WTP due to degree of freedom limitations. An R-squared of 0.77 is obtained but the only three significant coefficients are those for ICF , within design(at the 10% level) and calibration. Within design possibly gives IRSF closer to 1 either because of carry over effects or simply since the hypothetical and the real treatment groups are identical. This finding is consistent with evidence of a larger variance in responses in between subject designs in the experiments (Louviere, 2001). The calibration variable refers to either ex ante calibration techniques such as budget reminder or cheap talk scripts or ex post calibration such as using lab experiments to calibrate field data or uncertainty adjustments (Murphy et al., 2005). Likely the calibration techniques mitigates the erratic behaviour observed in hypothetical treatments. However, as in the case of the CF, for the SF we also lack a comprehensive theory that explains hypothetical bias. So, the causality of significant parameters should be interpreted with caution.

5. Conclusions

Our point of departure is the observation that most meta-analyses on discrete choice contingent valuation studies comparing real and hypothetical choice settings ignore the role of scale factor. Yet, efficiency question (bias and mean square error) is likely to be as important as the mere bias question in comparing stated and revealed preferences.

Building on Murphy et al (2005) our study provides some insights on the distribution of the inverse relative scale factor across 25 stated preference studies. The results show that, on average, the IRSF is about 0.6-0.7 and is correlated with the ratio between real and hypothetical average WTPs. However there are important differences in the distribution of the IRSF depending on which type of WTP distribution is considered: marginal wtp distribution, wtp model error distribution and wtp estimate distribution considered as non linear function of model parameters distribution.

6. Aknowledgments

The authors would like to thank, with the usual disclaimer, James Murphy for the comments provided on various versions of this paper.

Notes

---

[1] The authors would like to thank, with the usual disclaimer, James Murphy for the comments provided on various versions of this paper.

[2] Both dataset, bibliography and description of variable have been made available by Murphy on its own webpage at: http://faculty.cbpp.uaa.alaska.edu/jmurphy/meta/meta.html

[3] The five excluded studies are: Blumenschein, et al. (2001); Boyce, et al. (1989) ; Duffield and Patterson (1992); Murphy, et al. (2002) and Sinden (1988).

[4] A single study that directly provides the scale factor for a multinomial logit model also belongs to this group.

[5] A normal kernel density estimate was employed. The bandwidth parameter was selected with Sheater and Jones (1991) formula.

6. References

Blumenschein, K., Johannesson, M., Yokoyama, K., Freemand, P. (2001). Hypothetical versus real willingness to pay in the health care sector: results from a field experiment. *Journal of Health Economics* 20: 441-457

Boyce, R., Brown, T., McClelland, G., Peterson, G., Schulze, W. (1989). Experimental Evidence of Existence Values in Payment and Compensation Contexts. In: *Proceedings of the Joint Meetings of the Western Committee on Benefits and Costs of Natural Resource Planning* (W-133) , Western Regional Science Association, pp. 305-336.

Cameron, T. (2002). Alternative Non-market Value-Elicitation Methods: Are the Underlying Preferences the Same?. *Journal of Environmental Economics and Management* 44: 391-425

Cameron T., and James D.(1987). Estimating Willingness to Pay from Survey Data - an Alternative Pre-Test-Market Evaluation Procedure, *Journal of Marketing Research,* 24 (4), pp. 389-395.

Carlsson, F. and P. Martinsson (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments? - Application to the valuation of the environment, *Journal of Environmental Economics and Management*, 41, 179-192.

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

Diamond, P., Hausman, J. (1993) On contingent valuation measurement of nonuse values. In: Hausman, J.A., (ed) *Contingent valuation: A Critical Assessment*. North-Holland, Amsterdam, The Netherlands, pp 3-38

Duffield, J., Patterson, D. (1992). *Field Testing Existence Values: Comparison of Hypothetical and Cash Transaction Values*. Presented at Joint Western Regional Science Association/W-133. South lake Tahoe, Nevada .

Green, D., Jacowitz, K., Kahneman, D., McFadden, D. (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics* 20(2): 85-116

Haab, T., Huang, J., Whitehead, J. (1999). Are Hypothetical Referenda Incentive Compatible?: A Comment. *Journal of Political Economy* 107(1): 186-196

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

Hensher, D., Louviere, J., Swait, J. (1999). Combining sources of preference data. *Journal of Econometrics* 89: 197-221

List, J., Gallett, C. (2001). What Experimental Protocol Inßuence Disparities Between Actual and Hypothetical Stated Values? Evidence from a Meta-Analysis. *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. *Economics Bulletin*, 3(6): 1−13

Louviere, J. (2001) What If Consumer Experiments Impact Variances as well as Means? Response Variability as a Behavioral Phenomenon, *Journal of Consumer Research* 28(3): 506-511

Louviere, J., Hensher, D., Swait, J., Adamowicz, W.(2000) *Stated Choice Methods: Analysis and Applications*. Canbridge University Press, Cambridge (UK)

Murphy, J., Allen, G., Stevens, T., Weatherhead, D. (2005) A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation. *Environmental and Resource Economics* 30: 313-325

Murphy, J., Weatherhead, D. (2002) *An Empirical Study of Hypothetical Bias in Voluntary Contribution Contingent Valuation: Does Cheap Talk Matter?*. Paper prepared for the World Congress of Environmental and Resource Economists Monterey, CA .

Sheater, S., Jones, M. (1991) A Reliable Data-based Bandwith Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society* 53: 683-690

Sinden, J.A. (1988) Empirical Tests of Hypothetical Bias in Consumers' Surplus Survey. *Australian Journal of Agricultural Economics* 32: 98-112