Paul J. Metcalfe and William Baker

The sensitivity of willingness to pay to an economic downturn

Conference paper

The Sensitivity of Willingness to Pay to an Economic Downturn

Paul J. Metcalfe[1]*

William Baker[2]

Draft March 2012

1*Contact author: Paul J. Metcalfe, Department of Geography and Environment, London School of Economics, Houghton St, London WC2A 2AE, UK. Telephone: +44 (0) 1202 922712, Email: p.j.metcalfe@lse.ac.uk.

2. NERA Economic Consulting, 15 Stratford Place, London W1C 1BE, UK

Abstract

Stated preference (SP) studies are typically undertaken at one point in time, while the results may be relied on in decision-making several months or even years later. This reliance is only justified if values are stable over time, an assumption which is doubtable given the onset of an economic downturn. We assess the reliability of values taken before an economic downturn for application during the downturn, via analysis of responses to two near identical surveys conducted respectively before and during the 2008-2010 economic recession. The surveys were valuing near identical sets of permanent water sector service and environmental improvements. Each survey employed a dichotomous choice and a payment card contingent valuation question. Our main result is that the economic downturn led to lower payment card responses but had no effect on the values elicited via a dichotomous choice (ie referendum-type) contingent valuation question. We explore potential explanations for this finding in light of the literature on closed-ended versus open-ended elicitation method comparisons.

Keywords: Contingent Valuation, Stated Preference, Willingness to Pay, Water, Utilities, Regulation

JEL Classification: Q25, Q51, D60

1. Introduction

Stated preference (SP) studies are typically undertaken at one point in time, while the results may be relied on in decision-making several months or even years later. This reliance is only justified if values are stable over time, or are predictably different based on observable covariates. Fortunately, the weight of evidence suggests that this is often the case. A number of studies have administered similar questionnaires to independent samples at two points in time, and found that the estimated values, or valuation function, remained unchanged [*Brouwer*, 2006; *Brouwer and Bateman*, 2005; *Carson and Mitchell*, 1993; *Carson et al.*, 1997; *Reiling et al.*, 1990; *Whitehead and Hoban*, 1999]; a second group of papers have performed a repeated survey on the same sample of respondents, and found reasonably high correlations between responses [*Kealy et al.*, 1990; *Loomis*, 1990; *McConnell et al.*, 1998]. With one or two exceptions, the literature thus lends support to the application of values derived from historic contingent valuation surveys provided that reasonable adjustments are made for changes in observed determinants over the intervening period [*Whitehead and Hoban*, 1999].

There has been no study to date, however, which assesses the reliability of SP values obtained before an economic downturn for application during the downturn. There are theoretical and common-sense reasons to question whether willingness to pay (WTP) values, for example for environmental improvement, remain valid following the onset of a recession. Even after controlling for covariates such as current income, harder to observe potential explanatory factors such as perceived job security may be diminished, raising the possibility that willingness to contribute to the environment and related policy areas falls down the list of household priorities as a consequence. It is an open question whether the factors arising in a recession do indeed cause WTP values to fall, yet the answer has important implications for a wide range of policy applications.

The policy context in which the present study is situated is one such example. We conducted two near-identical contingent valuation surveys a year apart. The surveys were made of the household customers of a large English water and sewerage company as part of the five-yearly regulatory price review process, the first survey administered in June 2008 before the economic

downturn, and the repeat survey made on a new sample conducted in June 2009, when the UK was deep in recession. Each survey included payment card (PC) and dichotomous choice (DC) contingent valuation (CV) methods to elicit WTP values. The data from these two surveys thus provide the opportunity to test and compare the sensitivities of both PC and DC WTP responses to an economic downturn. Only one previous study [*Loomis*, 1990] has assessed the comparative reliability of these alternative elicitation methods; thus this feature of the paper makes an additional contribution to the literature by providing this comparison in an important new context.

2. A Model to Assess Temporal Sensitivity of WTP

Willingness to pay is typically specified as a function of observed covariates. Partly, this is to demonstrate that WTP varies in line with expectation; partly it is to allow for a more accurate transfer of values from one site and/or time period to another. In the following, to lay out the framework in which we consider the sensitivity of WTP to an economic downturn, we focus on the distinction between observed and unobserved WTP covariates, ignoring the features of the good and study site as these stay the same.

Let WTP for individual *i* in time *t* be written as:

$$WTP_{it} = f(\boldsymbol{x}_{it}, \boldsymbol{\psi}_{it}; \boldsymbol{\zeta}) \quad (1)$$

where $\boldsymbol{x}_{it}$ is a vector of observed covariates, $\boldsymbol{\psi}_{it}$ is a vector of unobserved covariates and $\boldsymbol{\zeta} = [\boldsymbol{\omega}^{\boldsymbol{x}}, \boldsymbol{\omega}^{\boldsymbol{\psi}}]$ is a vector of parameters. Note that $\boldsymbol{\zeta}$ is stable, that is, independent of *i* and *t*; all the variation over individuals and over time is captured by the two sets of variables $\boldsymbol{X}_t$ and $\boldsymbol{\Psi}_t$, where $\boldsymbol{X}_t = [\boldsymbol{x}_{1t}, \dots, \boldsymbol{x}_{Nt}]$, and $\boldsymbol{\Psi}_t = [\boldsymbol{\psi}_{1t}, \dots, \boldsymbol{\psi}_{Nt}]$.

Since $\boldsymbol{\psi}_{it}$ is unobserved, the following model is used as an approximation for estimation:

$$WTP_{it} = g(\boldsymbol{x}_{it}; \boldsymbol{\theta}_t) + \varepsilon_{it} \quad (2)$$

In (2), the unobserved covariates are no longer part of a deterministic function, and instead are captured by an error term, ε_{it}. Correspondingly, the functional form is changed from *f*(.) to *g*(.), and the associated parameter vector changes from $\boldsymbol{\zeta}$ to $\boldsymbol{\theta}_t$.

Estimation in time *t* typically relies on the identifying assumption that $E_i(\varepsilon_{it}|\boldsymbol{x}_{it}) = 0$. This is the case, for instance, when using OLS, tobit, logit, probit, or interval models, which are those most commonly employed to estimate valuation functions. The identifying assumption is generally invalid, however, if $\boldsymbol{X}_t$ and $\boldsymbol{\Psi}_t$ are correlated. Any correlation between the observed and unobserved covariates of WTP will cause the parameter vector $\boldsymbol{\theta}_t$ to be biased. Moreover, since the size of the coefficient bias depends on the unobserved data, and since this varies from year to year, the bias will itself vary from year to year. Only if the coefficients are unbiased, or if there is no substantial change in unobserved covariates, will the parameter vector stay stable from year to year.

In line with the terminology above, we assess the temporal reliability of WTP via the testing of two hypotheses:

(H1) $E\left(\overline{WTP}_2 \middle| \boldsymbol{X}_2, \widehat{\boldsymbol{\theta}}_1\right) = E\left(\overline{WTP}_2 \middle| \boldsymbol{X}_2, \widehat{\boldsymbol{\theta}}_2\right)$, and

(H2) $E(\boldsymbol{\theta}_2|\boldsymbol{X}_2) = E(\boldsymbol{\theta}_1|\boldsymbol{X}_2)$

The first of these hypotheses states that average WTP is predictable given new data on observed covariates of WTP, but using a previously estimated model. The second hypothesis makes the stronger claim which is that the predictive model is stable over time. Given estimates of $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_2$, these hypotheses may be straightforwardly tested by standard statistical methods. In section 4 we discuss estimation methods. We discuss the tests employed and their results in section 5.

3. Survey Design, Administration and Data

Thames Water (TW) is the largest water and wastewater services company in the UK supplying 8.8 million water customers and 14 million wastewater customers in London and the South East of England. In June 2008, we implemented a survey to assess its household customers' WTP for the

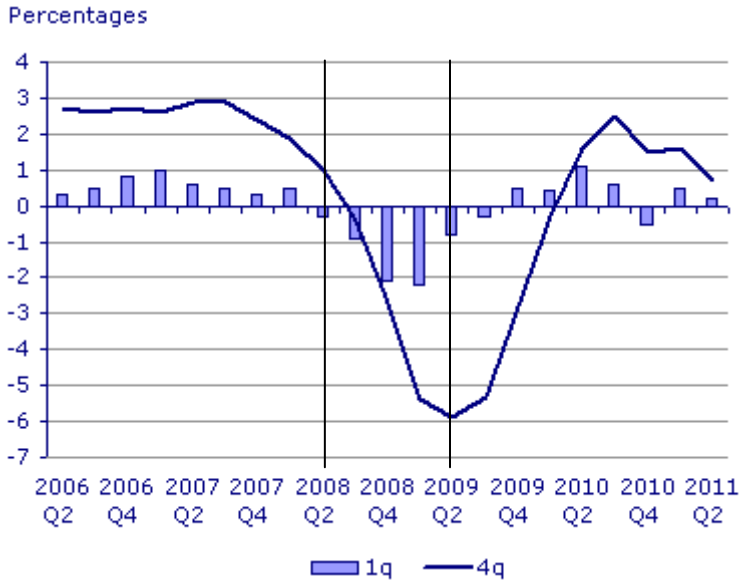
permanent improvements in water and wastewater service proposed in TW's draft business plan for 2010-2015, and in June 2009 we used a very similar questionnaire to assess household customers' WTP for the slightly revised set of permanent improvements set out in TW's final business plan. Both plans were submitted to the economic regulator Ofwat as part of its five-yearly price review process for the England and Wales water sector. Our analysis suggests that customers are likely to view the two sets of service improvements as being of extremely similar size, so from here on we refer to them both as simply "TW's plan". The appendix to this paper contains a table showing the details of current service levels (as also stated in the 2008 and 2009 surveys) and the list of service levels offered in the surveys and representing 2008 and 2009 investment plans.[1] The recruitment method, introductory questions, valuation statement and elicitation methods were the same in both surveys.

The dates over which effects should be most effectively measured can be debated. For example, in June 2008, although a recession had not yet been declared, there were already some warning signs of economic troubles ahead which could have influenced WTP responses at the time. On the other hand, in June 2009 unemployment had not yet reached its peak and so there is also a case to be argued that this later date may not capture the full impact of the recession on WTP. Both arguments would tend to suggest that any effect we estimate, in respect of the sensitivity of WTP to an economic downturn, is a lower bound. Ideally, for the purposes of this research, a series of surveys would be conducted to track changes in WTP over the full course of the economic cycle, a prospect which was unfortunately not feasible. As Figure 1 shows, however, the survey dates are situated at sufficiently different points in the economic cycle to have a good chance of capturing the effects we seek to examine.

Our method of examining separate samples has the advantage over a repeated survey on the same sample in that it eliminates any potential for recall bias, wherein the respondent remembers his

[1] We calculated the difference between the draft business plan (DBP) and final business plan (FBP) service improvement measures for each attribute, and used these to derive an index for the FBP based on the DBP and the current service level. If all proportional attribute improvements were given equal weighting by respondents, this approach determines that the FBP would imply "1% more" improvement than the DBP; i.e. probably a fairly trivial difference from the perception of respondents. Ideally we would use weights which match the relative values of the attributes rather than constant weights; however determining these weights is beyond the scope of this study.

original responses and simply repeats his answers in the second survey. A disadvantage is that only differences in population statistics, e.g. mean and median, can be compared, as we have no individual comparisons. Since population statistics are usually all that are needed for policy applications, and since these can be compared robustly using standard statistical methods, we do not consider this a significant limitation.



Note: vertical lines represent survey dates

Figure 1: UK Gross Domestic Product Growth, 2006-2011

The questionnaires each included a dichotomous choice (DC) contingent valuation question followed by a payment card (PC) question to elicit WTP for TW's plan. The payment vehicle was the annual water and wastewater bill increase; the levels for the DC question were drawn from the range {£5, £10, £20, £50, £100}; the payment card contained 30 numbers ranging from £0-£3000 on an approximately logarithmic scale. Many studies have found that DC values exceed those obtained by open-ended formats such as the PC approach [*Cameron et al.*, 2002; *Welsh and Poe*, 1998] to the extent that this is considered a 'stylized fact' of the CV approach [*Carson and Groves*, 2007]. [*Loomis*, 1990] is the only previous study, however, to have compared empirically the intertemporal reliability of alternative elicitation methods. It resurveys the same sample nine months after the original survey, asking DC and open-ended (OE) CV questions on each occasion, and finds the correlation between responses to be around 0.6 for both elicitation methods. Given the similarity of

OE and PC formats, we take this result as our prior that, in the absence of any wider change in conditions, we would expect PC and DC to be equally sensitive or insensitive to an economic downturn. We test this assumption as part of our analysis.

The surveys in 2008 and 2009 were administered face-to-face by Accent Market Research using the Computer Aided Personal Interview (CAPI) method. Each survey sought 300 responses stratified to include representative proportions of respondents in London, in rural areas, and in urban areas outside of London, with an average of 20 interviews per sampled location to ensure a dispersed sample. The average interview time was less than 30 minutes, and very few interviews took more than 40 minutes. The interviewers' comments on and scoring of respondents suggest that they understood the survey well, maintained a good degree of focus, and gave the questions careful consideration. Almost universally the respondents replied to a follow up question by stating that the cost, and/or the value to them of the service improvements, was the reason for their WTP answers. A fairly low proportion of the sample (9%) were excluded due to giving inadmissible responses to either the DC or PC questions. This comprised a mix of protest cases, refusals or "don't know" responses. A further 13% of the sample were excluded due to their failing to answer the income question. The final analysed sample sizes are 257 for the 2008 survey and 275 for the 2009 survey.

A summary of the respondent characteristics in the 2008 and 2009 surveys is presented in Table 1, alongside indicative population counterparts. The samples are broadly comparable, although the 2009 sample is somewhat older, better educated, higher earning and less likely to be a member of an environmental club. In respect of environmental club membership, this may be due to a decline in membership in the population rather than differences in sample composition – we are unable to confirm this either way. Population values in most cases are unlikely to be fully reliable due to the length of time since the UK census was conducted (2001). The exception to this rule is the case of income data for the London and South East region, which are drawn from the annual Family Resources Survey (FRS) for the relevant years. Based on a large-scale UK government survey, the FRS data offer a reliable picture of how household finances changed in the UK between 2008 and 2009. As Table 1 shows, nominal earnings appear to have risen slightly, despite the onset of a

recession. This is not altogether surprising since earnings, and employment, tend to lag behind output in the economic cycle. The small positive shift in the income distribution is reflected in the difference between the 2008 and 2009 samples that we obtained, however overall there are more low income respondents in our sample than in the population, and correspondingly fewer earning high incomes. To correct for this we adjust the sample observations with weights so that the analytical results reflect the income distribution of the population of household customers. This also ensures that the difference in income between the two samples, when weighted, matches the difference in income for the population. For our analysis, we also deflate 2009 income data, PC WTP and DC cost levels to 2008 prices using the Consumer Price Index (CPI) in order that the data and all reported results are comparable in real terms.

Table 1: Sample and Population Characteristics

	Population[(1)] (%)	2008 Sample (%)	2009 Sample (%)
Gender [(2)]			
Male	48.6	48.3	50.2
Female	51.4	51.8	49.8
Age [(2)]			
18-29	21.6	23.7	20.4
30-44	31.0	35.0	35.6
45-59	23.0	27.6	21.1
60-64	5.7	5.5	7.6
65-74	9.7	5.5	10.2
75+	9.0	2.7	5.1
Education [(2)]			
Primary	25.4	14.5	12.3
1-5 GCSEs/O-levels	16.1	25.4	18.4
5+ GCSEs/O-levels	20.5	13.3	15.7
2+ A-levels or NVQ3	10.1	15.3	17.6
First degree or higher	27.9	31.5	36.0
Employment Status [(3)]			
Working full-time (31+ hours)	42.9	47.6	46.0
Working part-time (<30 hours)	10.5	14.4	15.3
Self employed	9.3	4.8	7.7
Working and full-time student	2.8	1.6	2.3
Not working – seeking work	3.3	1.2	3.8
Not working – Full time student	5.3	4.4	2.7
Not working – retired	11.7	8.0	13.4
Not working – looking after home/family	6.8	10.8	6.9
Other	7.3	7.2	1.9
Weekly household income [(4)]			
Low (<£300)	22.3; 20.0	42.8	38.2
Medium (£300-£1000)	50.3; 52.4	45.1	44.4
High (>£1000)	27.4; 27.9	12.1	17.5
Environmental club membership [(5)]		19.8	15.3

Notes: N = 257 (2008 survey); N=275 (2009 survey). Base for each statistic includes the full sample unless indicated otherwise. (1) All population statistics are for the London and South East Government Office Regions combined. This region encompasses, and is somewhat broader than, the Thames Water supply area. (2) Source: Census (2001); (3) Source: Census (2001) (population aged between 16 and 74); (4) Source: Family Resource Survey (FRS); the first number in each pair is sourced from FRS (2008-09), representing the 12 months to March 2009; the second number in each pair is sourced from FRS (2009-10), representing the 12 months to March 2010; no adjustments have been made for inflation or other factors. (5) No population statistics available for environmental club membership in the region.

4. Empirical Methods

We analyze the data obtained from the survey as follows. First we combine the DC and PC responses using a single estimation technique - interval censored regression - and estimate this separately using the 2008 and 2009 samples. Interval frameworks are well suited to representing both DC and PC responses. [*Cameron and Huppert*, 1989; 1991] have argued that the language of a payment card question lends itself to an interval interpretation, with WTP lying between the amount indicated and the next highest amount labeled on the card. Interval frameworks have also long been used to

represent DC responses [*Carson and Hanemann*, 2005] with a no response indicating that WTP lies between zero and the amount asked and a yes response indicating that WTP lies between the amount asked and an upper bound reflecting financial resources. To be conservative, we use an upper bound of £500 for the interval when a respondent said yes to the DC question, which is substantially higher than the largest amount used (£100). This does not rule out the possibility that larger WTP values are held by respondents, only that they were not observed in either our PC or DC data.

The interval censored framework is straightforward to implement in a maximum likelihood context. Let y_n be our interval censored variable, which we model as a linear function of explanatory variables x_n plus an i.i.d. error term ε_n with mean zero and variance σ^2. Then we have:

$$\text{Prob}(y_n) = F\left(\frac{y_n^U - \boldsymbol{x_n\beta}}{\sigma}\right) - F\left(\frac{y_n^L - \boldsymbol{x_n\beta}}{\sigma}\right) \quad (3)$$

which implies the following log-likelihood:

$$\text{LL} = \sum_n \log[\text{Prob}(y_n)] \quad (4)$$

A distributional assumption is required for $F(.)$ to implement the estimation. We chose the log-normal because it ensures that WTP is non-negative (a problem with the normal) and it is straightforward to implement. Since the lower bound for some intervals is zero, the number "1" was added to all lower and upper bound values before taking logs because the log of zero is undefined. This "1" was then subtracted in obtaining later estimates for mean and median WTP. In the panel context, where for each person, n, we have a PC and a DC response, indexed by t, we thus let $y_{nt} = \log(1+WTP_{nt})$ and define lower and upper bounds accordingly, where WTP_{nt} is the willingness to pay by respondent n, as elicited by question type t ($t \in \{\text{PC, DC}\}$). $F(.)$ is then simply the standard normal cumulative distribution.

The log likelihood in (4) is based on the assumption that error terms are independent of one another. Independence is unlikely, however, when responses to both PC and DC questions are combined. To take account of within-person correlation between responses, we also estimate a random effects panel version of the above model which involves decomposing the error term into an individual specific effect, u_n, assumed to be normally distributed with mean zero and variance σ^2_u and an i.i.d. normal variate with mean zero and variance $\sigma^2_{e.}$ Estimation is performed using the *xtintreg*

command in *Stata* (version 11), and details of the methods and formulae can be found in *StataCorp* [2009].

5. Results

We begin by presenting the (weighted) response distributions for the PC and DC questions in 2008 and 2009 surveys. Consistent with the results of previous studies e.g. [*Welsh and Poe*, 1998], Figure 2 shows that the DC distribution lies above the PC distribution at all cost amounts for each year, except at the £5 level for the 2008 sample. Comparing across years, we see that the PC response distribution for 2009 lies below the 2008 distribution across the entire support, whereas for the DC responses there is no clear systematic difference. To examine this further we turn to presentation of our interval models, from which we can derive comparable estimates of mean and median WTP for PC and DC methods for 2008 and 2009, and the standard errors around these estimates that allow for statistical testing of the differences between them.

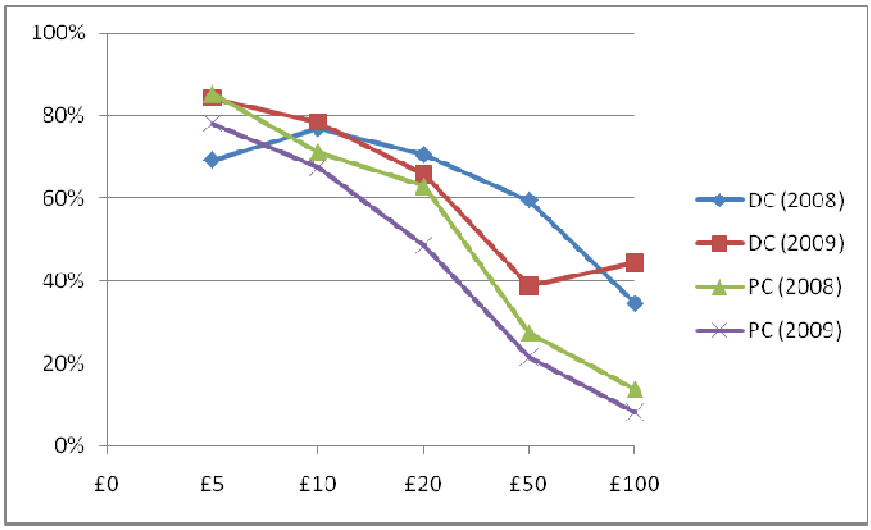


Figure 2: PC and DC Response Distributions in 2008 and 2009
Cumulative response frequencies offering at or above the WTP indicated amount, linearly interpolated between the DC levels used

Results from the interval models are presented in Table 2. The first model is for the 2008 survey sample. In this model, as anticipated, we see a significant ($p<.01$) negative coefficient on *Payment card*. The value of -0.381 indicates that PC WTP is around 32% lower than DC WTP all else equal. Turning to respondent covariates, income is positively associated with WTP ($p<.01$), again as expected. It enters in log form and so the coefficient on *Log income* is an elasticity; hence, the

coefficient of 0.509 implies that a 10% increase in income is associated with a 5% increase in WTP. Membership of an environmental club enters the model as a dummy variable, with a positive coefficient ($p<.05$), and via an interaction with Log income which has a negative coefficient ($p<05$). The combination of these two coefficients indicates that members of environmental clubs tended to have higher WTP than non-members except for the highest income respondents. The parameter σ_u is the standard deviation of the random effects. The fact that this is significant ($p<.01$) indicates that the random effects are themselves jointly significant. Consistent with this finding, the coefficient on ρ is 0.416, which indicates that 41.6% of the error variance is accounted for by the random effects. This evidence provides strong support for the use of the random effects interval model, rather than the simpler pooled model which assumes independence of individuals' errors across the two elicitation methods.

In comparison with the 2008 model, all the coefficients in the 2009 model seem very different, suggesting a lack of transferability of the full 2008 parameter vector for use during the recession of the following year. The coefficient on Payment card is -0.707 in the 2009 model which is lower than in the 2008 model. Whereas in the 2008 model, PC WTP is around 32% lower than DC WTP all else equal; in the 2009 model PC WTP is around 52% lower. The income elasticity is also much lower in the 2009 model than in 2008, at least for those that are not members of environmental clubs. The effect of club membership generally, as a function of income, is very different in 2009 than in 2008 which suggests that the original combined PC and DC function was not particularly reliable. The one aspect of the original combined function which does remain stable is the error distribution, as measured by σ_u, σ_e and ρ. Thus the shape of the distribution, if not its conditional means and medians, remains stable despite the onset of the economic downturn.

Table 2: Interval Censored Models Combining DCCV and PCCV Responses

Variable	2008[a,b,c,d]		2009[a,b,c,d]	
	Coef	Std Err	Coef	Std Err
Constant	0.351	(0.560)	2.583	(0.510)***
Payment card	-0.381	(0.101)***	-0.707	(0.098)***
Log income	0.509	(0.089)***	0.137	(0.081)*
Club	2.888	(1.334)**	0.120	(1.346)
*Club*Log income*	-0.436	(0.205)**	0.084	(0.202)
σ_u	0.754	(0.076)***	0.763	(0.073)***
σ_e	0.893	(0.056)***	0.894	(0.054)***
ρ	0.416		0.422	
Observations	514		550	
Log Likelihood	-841.042		-872.804	
Pseudo R^2	0.047		0.054	

Notes: ***a*** *Results are weighted for income based on the UK Family Resources Survey for the relevant year.* ***b*** *All models are interval censored regressions allowing for within person correlation. The left hand side for each model is the pair {*lry1,lry2*},where* lry1 *is the log of one plus the lower bound of WTP and* ly2 *is the log of one plus the upper bound of WTP, where WTP is measured in constant 2008 prices.* ***c*** *Standard errors are robust, calculated using the Huber-White estimator [*White*, 1980];* ***d*** *Stars indicate p-value for 2-side t test:* p < 0.10, ** p < 0.05 *** p < 0.01.*

As set out in section 2, we assess the temporal reliability of WTP via the testing of the following two hypotheses:

(H1) $E\left(\overline{WTP}_2 \middle| \boldsymbol{X}_2, \widehat{\boldsymbol{\theta}}_1\right) = E\left(\overline{WTP}_2 \middle| \boldsymbol{X}_2, \widehat{\boldsymbol{\theta}}_2\right)$, and

(H2) $E(\boldsymbol{\theta}_2 | \boldsymbol{X}_2) = E(\boldsymbol{\theta}_1 | \boldsymbol{X}_2)$

Hypothesis H1 states that predicted mean WTP in 2009 using new data ($\boldsymbol{X}_2$) but the original 2008 model ($\widehat{\boldsymbol{\theta}}_1$) is equal to our best estimate of actual mean WTP in 2009 based on both new data ($\boldsymbol{X}_2$) and a new model ($\widehat{\boldsymbol{\theta}}_2$). The second hypothesis, H2, makes the stronger claim that the predictive model is stable over time.

To test the stability of the 2008 valuation function (H2) we perform a Likelihood Ratio (LR) test to directly examine the suitability of the 2008 model coefficients for use in 2009. The 2009 equation presented in Table 2 is treated as the unrestricted model, and an equation also estimated on the 2009 sample but fixing all coefficients at the levels of the 2008 model, is treated as the restricted

model. This LR test rejects the null hypothesis of transferrable coefficients (p<.01), hence the combined PC and DC 2008 model is not transferrable to 2009. This finding is consistent with the readily seen differences between 2008 and 2009 models shown in Table 2.

The test of model stability is stronger than is usually necessary for cost-benefit analysis. In most cases, estimates of mean and median WTP are all that are required for policy applications. This is the motivation for the test of hypothesis (H1) – which states that predicted 2009 mean WTP from 2008 model coefficients is equal to predicted 2009 mean WTP from 2009 model coefficients. Given the functional form of the model, and letting $\widehat{\boldsymbol{\theta}}_t = \left[\widehat{\boldsymbol{\beta}}_t, \hat{\sigma}_t\right]$ where $\widehat{\boldsymbol{\beta}}_t$ is the vector of coefficient estimates for time *t* and $\hat{\sigma}_t$ is the estimate of $(\sigma_u + \sigma_v)$ for time *t*, we can write:

$$E\left[\overline{WTP}_t | \boldsymbol{X_t}, \widehat{\boldsymbol{\theta}}_t\right] = E_i\left[\left(WTP_{it} \middle| \boldsymbol{x}_{it}, \widehat{\boldsymbol{\theta}}_t\right)\right] = \exp\left(\widehat{\boldsymbol{\beta}}_t{}' \boldsymbol{X}_t + \frac{\hat{\sigma}_t^2}{2}\right) = g\left(\boldsymbol{X}_t, \widehat{\boldsymbol{\theta}}_t\right) \quad (5)$$

Then, following [*Whitehead and Hoban*, 1999], let the difference in WTP across time be

$$\Delta WTP_t = WTP_2 - WTP_1 \quad (6)$$

$$= g(\boldsymbol{X}_2, \boldsymbol{\theta}_2) - g(\boldsymbol{X}_1, \boldsymbol{\theta}_1) \quad (7)$$

$$= [g(\boldsymbol{X}_2, \boldsymbol{\theta}_2) - g(\boldsymbol{X}_2, \boldsymbol{\theta}_1)] + [g(\boldsymbol{X}_2, \boldsymbol{\theta}_1) - g(\boldsymbol{X}_1, \boldsymbol{\theta}_1)] \quad (8)$$

Table 3 presents this decomposition of WTP for the PC and DC predictions based on the estimated parameter vectors $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_2$, and the observed data $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Looking first at the PC results, the table shows that mean 2008 PC WTP - that is, predicted mean WTP using the 2008 model parameter vector and the 2008 data - was £46.1 per household per year. In 2009, mean PC WTP fell to £34.0. Changes in observable determinants ($\boldsymbol{X}$) caused a fall of £0.41 in PC WTP, although this difference is not significantly different from zero. The remaining £11.7 difference was caused by unobserved factors, and this difference is significantly different from zero (p<.01).

Now turning to the DC WTP results, we see that 2008 mean WTP was £68.0 per household per year, and in 2009 this rose to £69.9. Neither the difference attributable to changes in observed determinants (-£0.6), nor the difference attributable to changes in unobserved factors (£2.58) is

statistically significant (p>.10). The implication of these results is that PC WTP is sensitive to the economic downturn but DC WTP is not.

Table 3: Decomposition of Willingness to Pay, by Year of Data, Year of Estimated Parameter Vector and Elicitation Method

Year of Data (X)	Elicitation Method and Year of Estimated Parameter Vector ($\hat{\theta}$)			
	PCCV		DCCV	
	2008 ($\hat{\theta}_1$)	2009 ($\hat{\theta}_2$)	2008 ($\hat{\theta}_1$)	2009 ($\hat{\theta}_2$)
2008 (X_1)	46.1		68.0	
	(3.55)***		(5.20)***	
2009 (X_2)	45.7	34.0	67.4	69.9
	(3.40)***	(2.57)***	(4.98)***	(5.22)***
$g(\boldsymbol{X}_2, \hat{\boldsymbol{\theta}}_2) - g(\boldsymbol{X}_2, \hat{\boldsymbol{\theta}}_1)$	-11.7		2.58	
	4.27***		7.22	
$g(\boldsymbol{X}_2, \hat{\boldsymbol{\theta}}_1) - g(\boldsymbol{X}_1, \hat{\boldsymbol{\theta}}_1)$	-0.41		-0.60	
	4.92		7.20	

*Notes: Standard errors in parenthesis; standard errors are calculated using the delta method [*Greene*, 2003]; stars indicate p-value for 2-side t test:* p < 0.10, ** p < 0.05 *** p < 0.01.*

6. Discussion

The main findings are the following: (1) DC WTP is significantly higher than PC WTP in both survey samples; (2) the combined PC and DC valuation function as a whole was not found to be transferrable from 2008 to 2009; and (3) the onset of an economic downturn caused PC WTP to fall, while DC WTP remained unchanged. Finding (1) is consistent with the majority of the large number of studies that have compared DC and PC responses, as summarised in [*Champ and Bishop*, 2006]. The second finding gives cause for concern in using a combined valuation function derived before a recession during a recession, but it is not a sufficient finding to warrant disregard of predicted population mean WTP. Indeed, finding (3) suggests that if you believe that DC methods are more likely to get at the truth of WTP, then it is valid to predict mean WTP using a pre-recession DC valuation function, or just transfer the DC mean itself. We therefore focus our discussion on the implications of finding (3).

In light of the framework set out in section 2, we can infer from the findings that unobserved features of the downturn affected the PC responses but not DC responses. The principal unobserved

features potentially affecting WTP are, we hypothesize, diminished job security and a less certain future income – current incomes are, we have seen, not substantially different between years in our sample. We may now explore the consistency of these factors and the observed finding of no change in DC WTP but a fall in PC WTP with explanations given in the literature concerning the "stylized fact" that PC and OE responses are typically lower, sometimes much lower, than estimates of WTP generated from DC responses.

A prominent view in the literature, e.g. [*Carson and Groves*, 2007], explains the observed PC<DC relationship with reference to strategic response considerations. It is argued that the DC method is compatible with truth-telling provided certain stringent auxiliary conditions are met, namely that the DC question is asked before any other elicitation question, that the survey is constructed so as to convey the idea that respondents' answers will have a consequential impact on policy, and that respondents believe the scenario as presented to them, including the scope of the improvements and the cost they, and others, will have to pay. All three of these properties hold for the present study, and so it may be argued from this perspective that the DC WTP estimate is the truth. By contrast, under plausible belief structures – such as that the go/no go policy decision rule depends on summing respondents' stated PC WTP amounts, and that an individual's stated WTP amount is weakly correlated with the amount they will be required to pay should the policy be implemented – the PC method provides an incentive for respondents to understate their true WTP, either to minimize the chance that the policy goes ahead – stating a WTP of £0 when the cost is expected to be greater than true WTP – or to minimize the expected payment, by stating a WTP of the expected cost - sometimes rationalized as a "fair amount" - when the cost is expected to be less than true WTP. Strategic considerations are thus predicted to cause respondents to understate their true WTP when PC methods offer them the opportunity to do so.

For this view of the PC response process to hold, to explain our empirical finding of lower PC responses there would need to have been some change in incentives, or there must be some feature of a recession that causes respondents to become more strategically minded. The former condition can only be true if expectations of the true cost of the investment program had changed. Since there is no

difference in the information given in the survey, it is unlikely that cost expectations could have changed between surveys. On the other hand, it is plausible that increased job/income insecurity might invoke a greater willingness to engage in strategic response behaviour. Unfortunately, however, we are not able to test this hypothesis with our dataset.

A different perspective suggests that the observed difference between PC and DC WTP is due to differences in the certainty of respondents about their true WTP when they answer the questions [*Ready et al.*, 2001]. This view is backed up by some literature, showing firstly that respondents are indeed less certain about their DC responses than they are about their PC and OE responses, and secondly that equalising the certainty levels resolves the discrepancy [*Ready et al.*, 2001; *Welsh and Poe*, 1998]. To be consistent with this perspective, there would need to be some feature of a recession that caused respondents to become less certain of their true WTP. This seems plausible to us, in that increasing job insecurity might readily diminish certainty over WTP. This could cause there to be a wider uncertainty range, with a lower level of "certain" willingness to pay, but with no different a level at the top end of the range where respondents are "not sure" whether they would pay or not. Our results are thus also potentially consistent with this alternative explanation of the DC-PC difference.

7. Conclusions

Our main finding is that the recession caused PC WTP to fall, whereas DC WTP stayed the same. This result is statistically robust, hence the finding is probably not due to sampling variation. The two alternative principal explanations for the common finding that DC WTP>PC WTP – strategic behaviour, or respondent uncertainty - are both potentially consistent with our findings, and hence we cannot say for certain that one or other of them led the recession to cause PC WTP to fall, while leaving DC WTP unchanged. Consequently, we cannot say for certain whether true WTP itself is sensitive to an economic downturn. Until future research addresses this uncertainty, researchers will be able to interpret our empirical finding in line with their own views on which theory correctly explains the DC>PC differential.

References

Brouwer, R. (2006), Do stated preference methods stand the test of time? A test of the stability of contingent values and models for health risks when facing an extreme event, *Ecological Economics*, *60*(2), 399-406.

Brouwer, R., and I. J. Bateman (2005), Temporal stability and transferability of models of willingness to pay for flood control and wetland conservation, *Water Resources Research*, *41*(3).

Cameron, T. A., and D. D. Huppert (1989), OLS versus ML estimation of non-market resource values with payment card interval data, *Journal of Environmental Economics and Management*, *17*(3), 230-246.

Cameron, T. A., and D. D. Huppert (1991), Referendum contingent valuation estimate - sensitivity to the assignment of offered values, *Journal of the American Statistical Association*, *86*(416), 910-918.

Cameron, T. A., G. L. Poe, R. G. Ethier, and W. D. Schulze (2002), Alternative non-market value-elicitation methods: Are the underlying preferences the same?, *Journal of Environmental Economics and Management*, *44*(3), 391-425.

Carson, R. T., and R. C. Mitchell (1993), The value of clean water - the public's willingness-to-pay for boatable, fishable, and swimmable quality water, *Water Resources Research*, *29*(7), 2445-2454.

Carson, R. T., and T. Groves (2007), Incentive and informational properties of preference questions, *Environmental & Resource Economics*, *37*(1), 181-210.

Carson, R. T., W. M. Hanemann, R. J. Kopp, J. A. Krosnick, R. C. Mitchell, S. Presser, P. A. Ruud, V. K. Smith, M. Conaway, and K. Martin (1997), Temporal reliability of estimates from contingent valuation, *Land Economics*, *73*(2), 151-163.

Champ, P. A., and R. C. Bishop (2006), Is willingness to pay for a public good sensitive to the elicitation format?, *Land Economics*, *82*(2), 162-173.

Greene, W. H. (2000), *Econometric analysis*, 4th edition, Prentice-Hall.

Kealy, M. J., M. Montgomery, and J. F. Dovidio (1990), Reliability and predictive-validity of contingent values - does the nature of the good matter?, *Journal of Environmental Economics and Management*, *19*(3), 244-263.

Loomis, J. B. (1990), Comparative Reliability of the Dichotomous Choice and Open-Ended Contingent Valuation Techniques, *Journal of Environmental Economics and Management*, *18*(1), 78-85.

McConnell, K. E., I. E. Strand, and S. Valdes (1998), Testing temporal reliability and carry-over effect: The role of correlated responses in test-retest reliability studies, *Environmental & Resource Economics*, *12*(3), 357-374.

Ready, R. C., S. Navrud, and W. R. Dubourg (2001), How do respondents with uncertain willingness to pay answer contingent valuation questions?, *Land Economics*, *77*(3), 315-326.

Reiling, S. D., K. J. Boyle, M. L. Phillips, and M. W. Anderson (1990), Temporal reliability of contingent values, *Land Economics*, *66*(2), 128-134.

Welsh, M. P., and G. L. Poe (1998), Elicitation effects in contingent valuation: Comparisons to a multiple bounded discrete choice approach, *Journal of Environmental Economics and Management*, *36*(2), 170-185.

White, H. (1980), A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, *48*(4), 817-838.

Whitehead, J. C., and T. J. Hoban (1999), Testing for temporal reliability in contingent valuation with time for changes in factors affecting demand, *Land Economics*, *75*(3), 453-465.

Appendix

Table A1: Water and Wastewater Service Levels by Scenario

	Current Service[a]	2008 Plan[b]	2009 Plan[c]
Tap Water Service and its Climate Change Impact			
Risk of severe water rationing (rota cuts and stand-pipes for up to 3 months)	Expected 1 in 87 years	Risk eliminated	Risk eliminated
Leakage from Thames Water pipes	25% of water lost[d] 27% of water lost[e]	20% of water lost (20% reduction)	24% of water lost (11% reduction)
Unplanned interruptions to water supply of greater than 6 hours	13,000 households have an interruption each year	10,000 households have an interruption each year (25% reduction)	9,000 households have an interruption each year (31% reduction)
Drinking water quality (complaints about taste, colour and smell)	1,600 complaints per year	1,500 complaints per year (6% reduction)	1,500 complaints per year (6% reduction)
Carbon dioxide emitted by Thames Water caused by tap water service	No change from current levels	Fall of 10% (out of total fall of 20%)	Fall of 10% (out of total fall of 20%)
Wastewater Service and its Impacts on River Water Quality and Climate Change			
Households affected by sewer flooding	2,300 households at risk each year[d] 1,620 households at risk each year[e]	1,700 households at risk each year (26 % reduction)	1,180 households at risk each year (27 % reduction)
Improved quality of rivers and estuaries	No improvements in river water quality	225 km of river has improved quality	368 km of river has improved quality
Households affected by smell from sewage treatment	23,000 households affected each year	7,000 households affected each year (83% reduction)	7,500 households affected each year (67% reduction)
Carbon dioxide emitted by Thames Water caused by wastewater service	No change from current levels	Fall of 10% (out of total fall of 20%)	Fall of 10% (out of total fall of 20%)

Notes: ***a*** *"Current" levels of service shown were the same in both 2008 and 2009 surveys, except where indicated with notes d and e.* ***b*** *This column shows the levels of the improvement plan shown in the 2008 survey.* ***c*** *This column shows the levels of the improvement plan shown in the 2009 survey.* ***d*** *Level shown in 2008 survey only.* ***e*** *Level shown in 2009 survey only.*