Calibration of Willingness-to-Accept

John A. List  
Department of Economics  
University of Central Florida  
Orlando, FL 32816-1400  
John.List@bus.ucf.edu

Jason F. Shogren  
Department of Economics and Finance  
University of Wyoming  
Laramie, WY 82071-3985  
jramses@uwyo.edu

20 May 1998

Abstract

This paper calibrates real and hypothetical WTA estimates elicited for consumer goods in a multi-unit, random nth price auction. Our results suggest that people understated their real WTA in the hypothetical regimes, framed both as demand and non-demand revealing. Calibration factors from the unconditional model are as large as 1.42, and conditional marginal estimates from models of panel data range from 1.07 to 1.09.

JEL: Q21, Q26
Key words: calibration, hypothetical values, WTA

* We thank the National Science Foundation for financial support.
Calibration of Willingness-to-Accept

1. Introduction

Evidence suggests people often mask their real preferences when asked a hypothetical valuation question—recent research has shown they frequently overstate their real willingness to pay (WTP) in hypothetical markets.1 This observation triggered a search for a «calibration» function to test for and correct systematic bias between intentions and actions.2 Much less attention, however, has been paid to whether people distort their real willingness to accept (WTA) compensation, and little is known about the in-sample calibration of WTA offers.3 This is somewhat surprising since the recent hypothetical valuation literature has been driven by compensatory natural resource damage assessment, which is closely tied to WTA measures of value (e.g., Exxon Valdez and the Prince William Sound).

This paper uses panel data from a lab valuation experiment to calibrate in-sample real and hypothetical willingness to accept (WTA) compensation to surrender holiday gifts. We address two questions: 1) do hypothetical WTA offers differ across demand revealing and non-demand revealing elicitation frames? No—hypothetical offers were unaffected by the framing of choice; which provides further support that differences found in hypothetical valuation exercises cannot be explained away by referring to respondents having difficulty in answering open-ended questions; and 2) do hypothetical and real WTA offers differ? Yes—conditional and unconditional estimates suggest that subjects understated their real WTA under hypothetical scenarios; unconditionally, the ratio of mean real to hypothetical offers is about 1.5, and conditionally, the marginal calibration inflator is 1.07 to 1.09 per hypothetical dollar.

2. Data, Hypotheses, and Empirical Methods

3 Between-sample comparisons of hypothetical and real WTA offers include, amongst others, Bishop and Heberlein (1979), Bishop et al. (1983), and Brookshire and Coursey (1987). We are unaware of any studies that use within-sample data to calibrate hypothetical and real WTA offers.
We use data from List and Shogren’s (1998b) multi-good auction. Here we briefly summarize the three-stage experimental design: (1) hypothetical open-ended survey; (2) hypothetical demand revealing auction; and (3) actual demand revealing auction. In Stage 1 (February 4, 1997) the monitor asked 46 subjects at the University of Central Florida to complete Waldfogel’s (1993) hypothetical open-ended survey. Each person was asked to state «the amount of cash such that you are indifferent between the gift and the cash» for each of their 1996 Christmas gifts. For example, a person who received ten gifts submitted ten hypothetical WTA offers to sell the gift to the monitor. In Stage 2 (February 6, 1997) the monitor introduced a variant of a Vickrey-style uniform price, sealed-bid auction (Vickrey, 1961)—the random nth price auction. We used the uniform price auction to provide an incentive for each subject to truthfully reveal his or her value for each of his or her gifts. Again, the auction WTA offers were hypothetical. In Stage 3 (February 11, 1997) the monitor ran the actual random nth price auction in which each subject submitted his or her real value for each gift. All subjects knew they might potentially have to sell the gift to the monitor if they made one of the nth lowest offers. After accounting for attrition, our final sample consisted of 244 gifts across 36 recipients.

We use these WTA Christmas gift data to examine two basic questions. First, does the institutional frame matter for hypothetical WTA offers? Previous lab work on WTP bids has found mixed evidence regarding whether the institutional frame matters to hypothetical bidders (e.g., Neill et al., 1994; Frykblom, 1997; Balistreri et al., 1998). Also, the limited work that compares hypothetical and actual WTA responses has been between-sample and has yielded largely mixed results (see, e.g., Bishop and Heberlein, 1990 and the cites therein). We address this question by using our in-sample data to compare the mean hypothetical WTA offers: $H_0^{HI}: WTA^{HOE} = WTA^{HR}$, where $HOE$ and $HR$ are

---

4 After the instructions were read, the monitor ran a candy bar pre-auction to give the subjects some experience with the random nth price auction (see, e.g., Melton et al., 1996).

5 The auction works as follows: (1) for each gift received, a subject states his or her total value to sell the gift; (2) all gifts from each subject are then pooled together to create the set of total available gifts; (3) all gifts are then rank-ordered from lowest to highest total value; (4) the monitor selects a random number uniformly distributed between 2 and 21 (21 was the most gifts received by a subject); and (5) the monitor then purchases the $(n-1)$ lowest total value gifts overall and pays the nth lowest total value for each gift. Truth-telling can be a non-unique Nash equilibrium in a multiple-unit, uniform price auction, but in general this need not be the case (see Forsythe and Isaac, 1982). Lab evidence suggests that nth price auctions do a reasonable job of empirical demand revelation on aggregate (Miller and Plott, 1995; Franciosi et al., 1993).

6 The random price was #4, and the monitor made arrangements with the sellers to purchase the three lowest total value gifts each priced at the fourth lowest total value. Three distinct subjects sold one gift each at a uniform price of $2.00. We purchased hosiery, a beer intake facilitator, and slippers.
the hypothetical open-ended and hypothetical random n\textsuperscript{th} offers.\textsuperscript{7}

Second, do actual and hypothetical WTA offers differ? Most in-sample comparisons have focused on WTP bidding behavior, and the results suggest a significant gap exists between intentions and actions (e.g., Cummings et al., 1992, Fox et al. 1998, List and Shogren 1998a, 1998b). The few out-of-sample WTA comparisons had mixed results. Wisconsin goose hunters’ overstated their actual WTA to sell goose-licenses; deer hunters’ in a sealed-bid auction understated their actual WTA (but not significantly) to sell deer-permits, whereas those in a dichotomous choice institution overstated their real WTA (see Bishop and Heberlein’s 1990 summary). Also, Coursey et al. (1987) observed that people overstate their actual WTA to experience a drop of the bitter-tasting sucrose octa-acetate. But, since hypothetical values were elicited with an open-ended survey and actual values were elicited through a Vickrey auction, one cannot be sure that the real commitment was the catalyst for the valuation change.

We use two hypotheses to address the potential hypothetical-real WTA gap. First, the unconditional calibration hypothesis considers whether mean real and hypothetical WTA offers differ, $H_{0}^{ROE}$: $WTA_{RR}^{R} = WTA_{HOE}^{R}$ and $H_{0}^{RHR}$: $WTA_{RR}^{R} = WTA_{HR}^{R}$, where $RR$ denotes real random n\textsuperscript{th} price auction. Second, the no-bias calibration hypothesis examines whether hypothetical offers are consistent with real economic commitments. Since we have panel data–multiple offers across numerous subjects–our regression approach to test the no-bias calibration hypothesis controls for unobserved individual attributes that cannot be accounted for in tests of sample means or pooled ordinary least squares regression models.\textsuperscript{8} Given that subjects received an unequal number of Christmas gifts, we use an unbalanced panel data model to estimate the WTA calibration functions:

$$A_{ig} = \alpha_i + \beta H_{ig} + \varepsilon_{ig}, \quad i = 1, 2, \ldots, N; \ g = 1, 2, \ldots, G$$

\textsuperscript{7} An unconditional test holds nothing constant across subjects or gifts.

\textsuperscript{8} Studies that gather only one data point for each individual (due to their out-of-sample characteristics) cannot adequately control for unmeasured heterogeneity.
where $A_{ig}$ represents WTA from the real random $n^{th}$ price auction for the $i^{th}$ subject’s $g^{th}$ Christmas gift, $H_{ig}$ is the hypothetical open-ended or hypothetical random $n^{th}$ price auction WTA for the $i^{th}$ subject’s $g^{th}$ gift, $\alpha_i$ are the estimated fixed or random effects, and $\varepsilon_{ig}$ is the contemporaneous error term.

We estimate equation (1) using pooled ordinary least squares (OLS), fixed effects panel data, and random effects panel data models.\(^9\) The no-bias calibration hypothesis, using both sets of hypothetical data as the independent variable, is $H_{i}^{\text{No-Bias}}$: $\alpha_i = 0; \beta = 1$, where $\alpha_i$ and $\beta$ are defined above. Rejecting the no-bias hypothesis suggests that hypothetical WTA offers are inconsistent with real economic commitments.

### 3. Results

Table 1 presents the summary statistics, the Wilcoxon non-parametric tests of the equivalency of the WTA distributions, and the unconditional estimates of the calibration function for each of the three valuation methods. Mean offers in the three designs range from $95.77 to $136.87 per gift. Median offers of $40 are identical across the hypothetical question modes, while the real question mode yields a median offer of $50—a 25% increase relative to either hypothetical institution.

First, we cannot reject the framing hypothesis, $H_{i}^\text{HI}$: $WTA_{\text{HOE}} = WTA_{\text{HR}}$ (see Table 1). Although mean offers differ, we cannot reject the hypothesis that the revealed values in the hypothetical survey were derived from the same parental population as the values from the hypothetical auction (see Table 1, Wilcoxon test, $W = 3281 \ (z = -0.46)$). This observation supports previous results that the institutional frame does not universally affect hypothetical valuations. Several WTP studies find similar results (e.g., Neill et al., 1994 and Frykblom, 1997). If it is meaningful to know which hypothetical mode best approximates real values, this result suggests the gap between intentions and actions cannot be explained away by saying that people had difficulty answering an

\(^9\) Both fixed and random effects models control for unmeasured heterogeneity that pooled OLS ignores. Random effects estimates of (1) yield coefficients that are not conditioned on unmeasured person effects, whereas fixed effects estimates yield coefficients conditioned on the unmeasured characteristics. Fixed effects estimates are inefficient since they only consider within-person variation. Yet, if the person effects are correlated with hypothetical WTA responses, random effects estimates are biased and inconsistent, while the within estimator remains unbiased and consistent. We test for unbiasedness and consistency using a Hausman (1978) test when comparing estimates from fixed and random effects models.
open-ended question.

Second, we reject the unconditional calibration hypotheses—hypothetical and real WTA offers differ significantly, not controlling for subject effects. Table 1 shows the mean WTA ratio and the median WTA ratio across designs. Hypothetical offers underestimate real offers—the ratio of hypothetical and real random $n^{th}$ price means is about 1.4. These differences are statistically different ($W = 9404 (z=-3.84); W = 9560 (z=-13.25)$) at the 1% significance level, implying revealed values in the actual auction are not derived from the same parental population as data from the two hypothetical treatments. These calibration factors exceed those observed in WTP experiments, which range between 0.1 to 1 (see List and Shogren, 1998a).

We also reject the no-bias calibration hypotheses—hypoithetical and real offers differ controlling for unobservable subject-bias. Table 2 presents panel data estimates for ordinary least squares, and fixed and random effects models using both hypothetical WTA institutions as the independent variable.\(^\text{10}\) We observe that the estimated coefficients on both hypothetical offers are significant at the 1% confidence level. In the hypothetical WTA open-ended survey, the point estimate in column 2 implies that a dollar increase in the hypothetical offer increases the real offer in the random $n^{th}$ price auction by $1.09$, which is significantly different from $1$ at the 95% level. We find a similar result in the hypothetical random $n^{th}$ price auction fixed effects specification (column 5). Point estimates imply that a $1$ increase in the hypothetical random $n^{th}$ price offer increases the real offer by $1.07$, which is significantly different from $1$ at the 90% level. Although these findings would not lead to rejection of the no-bias calibration hypothesis at the 99% level, heterogeneity tests (F-statistic in Table 2) reject the null hypothesis of homogeneity of unmeasured subject-specific effects at the 1% level for the fixed effects models. These tests suggest that individual-specific intercepts are significantly different from zero at the $p < 1\%$ confidence level, leading us to reject the no-bias calibration hypothesis. This

\(^{10}\) Hausman (1978) tests of the null hypothesis of zero correlation between hypothetical offers and the subject-effects suggest that the orthogonality assumption underlying the random effects estimates is violated in both sets of regression models. This outcome suggests that GLS yields inconsistent and potentially biased coefficient estimates. Nevertheless, similar implications are drawn from both model types. Finally, note that the relatively parsimonious regression models appear to be sufficiently rich in that the adjusted $R^2$ from both fixed-effects models is 0.80.
finding supports the conjecture that there is an individual-specific, systematic component in the error term that leads to bias in hypothetical responses (see, e.g., Mansfield, 1998, Andreoni, 1995, Herriges and Shogren, 1996).

Calibration functions seem to be individual-specific. If within-person bias exists that significantly affects the calibration function, the overall relationship between hypothetical and real offers should account for the individual-specific effects \( \alpha_i \) in the regression model. Recent studies have suggested that respondent’s characteristics and attitudes may partly determine whether, and by how much, a person’s hypothetical bids differ from his or her real bids (Mansfield, 1998). Given that the fixed effects component of the regression captures any time-invariant unobservable/observable subject factors, we perform an exploratory probe for any systematic pattern in the individual-specific effects. To effectively carry this task out, we use the estimated \( \alpha_i \) from equation (1) as the dependent variable in the linear regression model given by:

\[
\alpha_i = \phi + \beta X_i + \epsilon_i
\]  

(2)

where \( \alpha_i \) are the estimated fixed effects from (2), \( \phi \) and \( \beta \) are the estimated intercept and slope parameters, \( X_i \) are attributes hypothesized to influence the fixed effects, and \( \epsilon_i \) is an error term. Evidence from the psychology literature suggests that individual response strategies may be a function of personal characteristics (see Krosnick’s review, 1991). As such, we consider four measurable attributes in vector \( X_i \)—AGE, GENDER (= 1 male, 0 female), FAMILY INCOME, and the number of gifts, #GIFTS, that person \( i \) could sell in the actual auction. Table 3 presents the descriptive statistics.

Table 4 shows the estimates of equation (2). Both models are significant at conventional levels, and estimates are similar across specifications. Focusing on the open-ended survey (column 1), we see that the fixed effects are at least partially determined by demographic factors as a majority of coefficients are significantly different from zero at conventional levels. More specifically, we find that
a subject was more likely to understate a real offer if he or she was older, a man, or had received more
gifts. For example, the coefficient on \#GIFTS suggests for each extra gift, hypothetical offers needed to
be increased by $14.90, from equation (1). This finding suggests that hypothetical auctions do not
necessarily provide incentive for people to work through the cognitive processes to evaluate each and
every good seriously. Overall, the psychology results cited in Krosnick (1991) support our general
observations. Although intuition suggests people should play it safe by stating large WTA values when
their actions or preferences are not well thought out (e.g., Hoehn and Randall, 1987), our results suggest
otherwise.

4. Concluding comments

Most practitioners view WTA measures as the upper bound on value for incremental changes in
a good or service (e.g., Cummings et al., 1987). Our results show that hypothetical offers could
actually represent a lower limit on this upper bound—the average person understated his real WTA
offers to sell a market good. Unconditional hypothetical-real calibration factors were as large as 1.4,
and conditional marginal calibration factors ranged from 1.07 to 1.09. As Mansfield (1998, p. 680)
points out: «the power of the calibration model could be improved by a better understanding of how
individuals answer CV questions, including the traits or attitudes that inspire individuals to give more
or less accurate answers.» Although we have provided some initial findings pertaining to these issues,
to construct a robust estimation procedure, more research is needed to uncover the relationship between
intentions and actions. An important start would be to test for robustness across goods and context.
References


Table 1. Summary Statistics and Non-Parametric Tests of Equivalency

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Bid</td>
<td>$95.77</td>
<td>$96.34</td>
<td>$136.87</td>
<td>3281&lt;sup&gt;b&lt;/sup&gt; z = -0.46</td>
<td>9404* z = -3.84</td>
<td>9560* z = -13.25</td>
</tr>
<tr>
<td></td>
<td>($12.28)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>($12.48)</td>
<td>($15.36)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Bid</td>
<td>$40</td>
<td>$40</td>
<td>$50</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

Calibration Factors

<table>
<thead>
<tr>
<th></th>
<th>Mean WTA Ratio</th>
<th>Median WTA Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>WTA Ratio</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

N 244 244 244 --- --- ---

<sup>a</sup>Standard errors in parentheses.
<sup>b</sup>Wilcoxon test is a signed-rank test for matched pairs across gifts. Since the number of paired observations is larger than 30, the large-sample z test is used. The large sample z-test’s null and alternative hypotheses are given by: H<sub>0</sub>: Two sampled populations have identical probability distributions. H<sub>a</sub>: The probability distribution for population A is shifted to the right or to the left of that for population B. Critical z-values are computed as follows: z = (W - (n(n+1)/4))/(n(n+1)(2n+1)/24)<sup>1/2</sup>. Where n is the number of non-tied differences between the two samples.
<sup>c</sup>Ratios are calculated as the top value in the column header divided by the lower value in the column header. For example, mean bid ratio under the fifth column (labeled actual auction versus hypothetical auction), is computed as (mean actual auction/mean hypothetical auction).

* Significantly different values at the 1% level.
Table 2. Fixed and Random Effects Estimation Results for Calibration Functions\textsuperscript{a,b}

<table>
<thead>
<tr>
<th>Variable</th>
<th>\textit{Open-Ended Survey}</th>
<th></th>
<th>\textit{Hypothetical Auction}</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>Fixed</td>
<td>Random</td>
<td>OLS</td>
</tr>
<tr>
<td>\textit{Constant}</td>
<td>43.1*</td>
<td>---</td>
<td>24.6</td>
<td>44.4*</td>
</tr>
<tr>
<td></td>
<td>(4.0)</td>
<td>(1.2)</td>
<td>(10.8)</td>
<td>(10.8)</td>
</tr>
<tr>
<td>\textit{Hyp. Offer}</td>
<td>0.98</td>
<td>1.09*</td>
<td>1.06*</td>
<td>0.96*</td>
</tr>
<tr>
<td></td>
<td>(19.6)</td>
<td>(26.1)</td>
<td>(26.2)</td>
<td>(19.4)</td>
</tr>
<tr>
<td>\textit{R}^2</td>
<td>0.61</td>
<td>0.84</td>
<td>---</td>
<td>0.61</td>
</tr>
<tr>
<td>\textit{Adj. R}^2</td>
<td>0.61</td>
<td>0.80</td>
<td>---</td>
<td>0.61</td>
</tr>
<tr>
<td>\textit{F(\alpha = 0)} (d.f.)</td>
<td>---</td>
<td>7.80*</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(35, 207)</td>
<td></td>
<td>(35, 207)</td>
<td></td>
</tr>
<tr>
<td>\textit{LM(\alpha = 0)} (d.f.)</td>
<td>---</td>
<td>---</td>
<td>67.1*</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>\textit{Hausman} (d.f.)</td>
<td>---</td>
<td>---</td>
<td>8.9*</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>\textit{N}</td>
<td>244</td>
<td>244</td>
<td>244</td>
<td>244</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Dependent variable is actual offer in random nth price auction.
\textsuperscript{b}t-statistics in parentheses under coefficient estimates.
*Significant at the .01 level.
Table 3. Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FIXED EFFECTS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open-Ended Survey</td>
<td>17.65</td>
<td>139.2</td>
</tr>
<tr>
<td>Hypothetical Auction</td>
<td>21.16</td>
<td>140.0</td>
</tr>
<tr>
<td><strong>AGE</strong></td>
<td>24.22</td>
<td>7.0</td>
</tr>
<tr>
<td><strong>GENDER</strong></td>
<td>0.58</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>#GIFTS</strong></td>
<td>6.78</td>
<td>3.99</td>
</tr>
<tr>
<td><strong>FAMILY INCOME</strong></td>
<td>$68,472</td>
<td>$90,385</td>
</tr>
</tbody>
</table>

N 36

*Gender = 1 if male, 0 if female.*
Table 4. Determinants of the Fixed Effects Component of the Calibration Function\textsuperscript{a,b,c,d}

<table>
<thead>
<tr>
<th>Variable</th>
<th>Open-Ended Survey</th>
<th>Hypothetical Auction</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>-285.7** (-2.7)</td>
<td>-277.5** (-2.6)</td>
</tr>
<tr>
<td>AGE</td>
<td>6.3** (2.0)</td>
<td>6.1* (1.9)</td>
</tr>
<tr>
<td>GENDER</td>
<td>99.3** (2.0)</td>
<td>105.3** (2.1)</td>
</tr>
<tr>
<td>#GIFTS</td>
<td>14.9** (2.2)</td>
<td>14.3** (2.0)</td>
</tr>
<tr>
<td>FAMILY</td>
<td>-0.12 (-0.3)</td>
<td>-0.11 (-0.4)</td>
</tr>
<tr>
<td>INCOME</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>N</td>
<td>36</td>
<td>36</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Dependent variable is the fixed effect $\alpha_i$ from equation (2).
\textsuperscript{b}t-ratios are beneath coefficient estimates.
\textsuperscript{c}Family income coefficients are multiplied by 1000.
\textsuperscript{d}Gender = 1 if male, 0 if female.
** Significant at the .05 level.
* Significant at the .10 level.