

Calibration of the difference between actual and hypothetical valuations in a field experiment

John A. List^{a,b,*}, Jason F. Shogren^b

^a *Department of Economics, University of Central Florida, Orlando, FL 32816-1400, USA*

^b *Department of Economics and Finance, University of Wyoming, Laramie, WY 82071-3985, USA*

Received 6 January 1997; accepted 1 September 1997

Abstract

We design and implement a field experiment to elicit and calibrate in-sample hypothetical and actual bids given the presence of other goods and intensity of market experience. Using market goods that possess characteristics beyond the norm but yet remain deliverable, bidding behavior was consistent with theory. But we also observe the average calibration factor for hypothetical bids in the auction with other goods to be more severe (0.3) than for the auction without the goods (0.4). The results support the view that the calibration of hypothetical and actual bidding is good- and context-specific. © 1998 Elsevier Science B.V. All rights reserved.

1. Introduction

Understanding whether people overstate their actual preferences for a good when asked a hypothetical question remains an important issue in current policy debates over environmental goods (e.g. how much would you pay to protect the Wyoming toad?). Earlier work suggested that the average person exaggerates his actual willingness to pay (e.g. Bohm, 1972; Bishop and Heberlien, 1979; Dickie et al., 1987; Shogren, 1990; Seip and Strand, 1992; Neill et al., 1994; Carson et al., 1996). In response, the National Oceanic and Atmospheric Administration's (NOAA) blue-ribbon panel recommended that hypothetical bids be deflated using a 'divide by 2' rule unless these bids can be calibrated using actual market data (NOAA, 1994, 1996). The NOAA rule has served an ad hoc placeholder to motivate more research into the nature of calibrating hypothetical

* Corresponding author.

and actual values (e.g. the statistical bias functions in Blackburn et al., 1994 and the CVM-X design in Fox et al., 1998).¹

This paper implements a field experiment to further examine calibration by comparing bidding behavior in a hypothetical and actual second-price auction for baseball cards – deliverable objects with an intangible quality. Baseball cards have many favourable characteristics for a calibration exercise including familiarity, the ability to deliver, and an abstract quality beyond the normal market good.² We run one sample with one card, a second with one card among 10, and a third with one card bid on by sportscard dealers presumed more experienced with the market than the general population.

The results suggest that bidding behavior is consistent with theoretical expectations as the inclusion of other goods dampened mean hypothetical and actual bids, and market experience reduced bid variance. But the results also show that calibration is sensitive to the context of choice: (a) the presence of other goods – hypothetical bids required less deflation in the 1-good auction than the 10-good auction; and (b) market experience – hypothetical bids of dealers needed less deflation than non-dealers. While more stringent than other estimates (e.g. Fox et al., 1998), our evidence supports the view the calibration is good- and context-specific.

2. The good, the design, and the hypotheses

Our objective is to calibrate hypothetical and actual bids for 1-good with and without the presence of other goods, and for different intensities of market experience. We conducted three treatments at a sportscard show in Denver, CO in December 1995: the 1-good, 10-good, and Dealer treatments.

2.1. *The good*

For the 1-card and Dealer treatments, the auctioned good was a Cal Ripken Jr. 1982 Topps Traded PSA-graded 9 rookie baseball card.³ All treatments displayed the same Cal Ripken Jr. card to ensure comparability of bids. An independent agency, Professional Sports Authenticators (PSA), graded the Cal Ripken Jr. card to avoid complications of participants not understanding the grade (i.e. substance and quality) of the card. We

¹ Randall, 1997, p. 200) states that: “[t]he calibration issue, it seems to me, is an audacious attempt to promote a Kuhnian paradigm shift. ... I would argue vigorously that the essential premise is unproven and the question is, therefore, premature and presumptuous. The proposed new calibration paradigm is at this moment merely a rambunctious challenger to the dominant external validation paradigm.” One can interpret this statement as a call for more work on calibration.

² A baseball card is a piece of cardboard that serves as a proxy for more general preferences on the intangible aesthetics of indirect athletics. Without attempting to explain the psychology of sports fans, people often like to feel like they own a part of the team or moment: sportscards can help fulfill that preference.

³ Neill et al. (1994) compare hypothetical to actual values for goods with intangible qualities: watercolor paintings and a map. Several methodological differences exist between our study and theirs: they examine between-sample values, we consider in sample calibration; they do not address the impacts of other goods on value; and their experiments were in the lab with students, our auctions are in a familiar marketplace with actual collectors. Also, see Bohm’s (1984) work on revealing actual values for public goods in field experiments.

selected the Cal Ripken Jr. card to reduce the valuation issues encountered when subjects do not understand the substance of the good they are asked to value (Cummings et al., 1986). Most collectors are familiar with the 1982 ‘Topps Traded’ rookie card of Cal Ripken Jr.: the sportscard collectors’ bible, Beckett magazine, has recognized this card as a ‘hot list’ member for over 4 years with listed book value of \$350 (October 1995). Additionally, perusal of any sportscard show reveals Ripken’s popularity and familiarity with collectors, especially with his well-publicized record-setting 2,131+ consecutive games.

For the 10-good treatment, we used the Cal Ripken Jr. card plus nine other sportscards (or sets of cards) that could act as potential substitutes or complements. Two other Cal Ripken Jr. 1982 ‘rookie’ PSA-graded cards, 1982 Topps and 1982 Donruss, served as substitutes⁴. The choice of substitutes was motivated by the observation that many collectors view another rookie card of the same player as a substitute. Complements included the complete set of 1982 traded cards without Cal Ripken Jr.’s card and Billy Ripken’s 1989 rookie PSA-graded card. We selected complements on the basis of many people wanting to complete sets (e.g. 1982 Topps Traded without Ripken Jr.’s card); and many collectors wanting to complete ‘oddball’ sets, such as brothers, or double-play partners; a Billy Ripken card fits both criteria. Five other goods were included in the 10-card auction to provide additional substitutes; these were football cards consisting of three Troy Aikman cards, one complete set (without Aikman), and one Michael Irvin card. All cards were independently graded by PSA.

2.2. *Experimental design*

Our auction for the 1-good and 10-good treatments followed a four-step experimental design: (1) inspection of the good(s), (2) hypothetical bid(s), (3) actual bid(s), and (4) debriefing. In Step 1, monitor A approached a person entering the show and asked if he or she would like to participate in a hypothetical auction that would take about 10 minutes. If the individual agreed, the monitor briefly explained that we were hypothetically auctioning off the baseball card(s) displayed on the table. The participant could pick up and visually examine each card. All cards were sealed with the PSA grade clearly marked on each cardholder. The monitor worked one-on-one with the participant and no time limit was imposed on his or her inspection of the card(s). We did not give the participants any financial incentives or gifts to participate, thus, we avoid any claims of the results being influenced by ‘found money’ effects. After a participant had examined the card(s), Step 2 began.

In Step 2, monitor A gave the participant an instruction sheet that consisted of two parts: (i) a short socio-economic survey (e.g. age, education, years trading), and (ii) a bidding sheet (Experimental instructions are available on request.) The participant was asked to submit a hypothetical bid stating the maximum that he or she was willing to pay

⁴ The difference between the 1982 ‘Topps Traded’ and 1982 ‘Topps’ cards is that the ‘Traded’ card is an issue focusing on ‘rookies’ and players who are traded to a new team the previous season (or over the off-season), while the ‘Topps’ card is from a separate product line which includes the majority of players.

for the card(s). The instructions for the bidding sheet stated that the hypothetical exchange mechanism was a sealed bid second-price auction. Specifically, the bidding sheet reported:

A sealed bid *second-price auction* will be used to determine the winner of this item. Thus, if your bid of \$X for this item is the highest bid and the next highest bid is \$X-5, you win this item but will only pay \$X-5. Under this bidding mechanism it is best for you to bid your *true value* for this item because overbidding may cause you to pay too much and underbidding decreases your odds of winning the item.

Note: You will not be required to pay this amount and all bids are hypothetical. Also, the winner will not receive this card.

Following earlier experimental auctions (e.g. Coursey et al., 1987; Shogren et al., 1994), we use the second-price auction given its theoretical incentive compatibility properties (Vickrey, 1961). While not flawless, the second-price auction has performed reasonably well in revealing preferences for both induced and non-induced value auctions (see, e.g. Kagel's, 1995 guarded but positive evaluation). After the participant filled out the survey and hypothetical bidding sheet privately, he or she folded the bidding sheet and placed it in an opaque box. The monitor told the participant that his or her bid would not be opened until after show and that all bids would be destroyed when our research project was complete. Monitor A then asked the participant to go over to monitor B at a second table 15 feet away for a follow-up auction.

In Step 3, monitor B told the participant that he or she now had the chance to actually bid on the card(s) that he or she had just examined in Step 2. Monitor B gave the participant a second bidding sheet for the actual auction. Again we used the sealed bid second-price auction as the exchange mechanism. After the monitor answered all questions about the auction, the participant placed his or her sealed bid into a second opaque box. To guarantee that we did not get a second hypothetical bid monitor B asked each participant to acknowledge their actual bid with a signature and valid telephone number where they could be contacted. Care was taken to avoid contamination of the results by any ordering effects (e.g. sealed boxes, monitors not handling or observing bids). Cummings et al. (1995) found no evidence of an ordering effect in a dichotomous choice setting.

Finally, in Step 4 monitor B debriefed the participant. The monitor explained that the actual bids would be ranked, and the participant would be contacted within 3 days after the show if he or she was the highest bidder. Monitor B also explained that if the participant won the auction, he or she would receive the card(s) after he or she had sent a check or money order for the amount of the second highest bid. After any remaining questions were answered, monitor B asked the participant not to discuss the auction with anyone else until after the show, and then thanked him or her for participating in the project. Within 3 days the winner of each auction was notified by phone and when the monitors received the checks they mailed out the cards.

The 1-good and 10-good treatments took approximately 12 hours to complete (from 9 A.M. to 9 P.M.). On the top of each hour the auction treatment was switched from the 1-good to the 10-good treatment and vice versa the next hour. No participant took part in

both auctions. Participation rates were 82 percent (99 of 121) for the 1-good treatment, and 84 percent (93 of 111) for the 10-good treatment⁵.

The Dealer treatment was similar to the 1-good treatment except that a monitor visited each dealer at his or her booth the night before the sportscard show. The monitor first gave each dealer an instruction sheet for the hypothetical auction, and then administered the follow-up actual, upon the promise that they would not leak any information to potential subjects. The treatment took about 2 h (from 6:30 to 8:30 P.M.), and the participation rate was 91 percent (30 of 33).

2.3. Auction institution and economic hypotheses

Let (b_{Hk}^i, b_{Ak}^i) represent bidder i 's hypothetical and actual bids for the Cal Ripken, Jr. Topps Traded card, where subscript H and A represent the hypothetical or actual bid, $k=one$ or ten represents the 1-good or 10-good case, and the superscript i is the bidder ($i=1, 2, \dots, n$). In the 1-good case, a person's dominant strategy is to bid his or her true value in the second-price auction (Vickrey, 1961). In the 10-good case, the strategy is more involved since the bidder can win one or more goods. People in a simultaneous auction can be viewed as either solo bidders wanting 1-good, local bidders wanting some smaller set of goods, or global bidders who derive benefits from all the goods (Krishna and Rosenthal, 1995). A solo, local, or global bidder's bid should equal his or her true value for each good, depending on whether he or she wants 1-good or some combination. Given a binding budget constraint, a person allocates his or her financial resources such that bids satisfy standard marginal efficiency conditions – the ratios of expected marginal benefits and costs should be equated across goods. For more on multiple-good auctions, see, McMillan (1994); McAfee and McMillan (1996); Melton et al. (1996).

Given these two institutions, we compare the hypothetical and actual market price bids, b_{jone}^2 and b_{jten}^2 $j = H, A$ (i.e. the second highest price), to Beckett's October 1995 book value for the cards. We estimate the percentage of market book value received in the hypothetical and actual auctions. If the bids are not within a reasonable range of a market benchmark, the bidders might not take the exercise seriously, and the results could be viewed as problematic. We define a 'reasonable' range as within 200 percent of book value; other definitions can be easily accommodated.

We also consider two measures of internal consistency regarding the bids from our auctions. First, we test how the availability of other goods affect hypothetical and actual

⁵ In the lab, more than one trial is often required before people understand the nature of the second-price auction. We made a pragmatic choice to use a one-shot auction so that we could run the experiment on the floor of the sportshow. To test whether subjects understood the second-price auction quickly, we ran a pilot study in November 1995 at a similar sportscard show in Denver using the instruction sheet and experimental procedures. On completion of the survey and bidding sheet, subjects were questioned about their understanding of the good and the second-price auction. Nineteen people took part in the pilot study, and no one had any problem understanding the auction, questions, or provisioning rule. Also a one-shot auction removes the potential for upward bias stimulated by new information in an auction with affiliated common values (Milgrom and Weber, 1982). Baseball cards, especially for the dealers, can have a substantial common value component given the going market price is well-publicized.

bidding behavior. Bids should decline because values decline with more substitution possibilities, holding the binding budget constraint constant (i.e. the Le Chatelier principle). Valuation research has supported this prediction for hypothetical goods (e.g. Cummings et al., 1994). In our case, we should reject the hypothesis of identical populations of Cal Ripken, Jr. bids in the 1-good and 10-good auctions. The central tendencies of hypothetical and actual bids for the Cal Ripken, Jr. card in the 1-good auction should exceed the mean bids in the 10-good auction, $\bar{b}_{jone} > \bar{b}_{jten} (j = H, A)$.

A second consistency check is that dealer bids – those with more intense experience with market prices – will be more clustered than no-dealer bids. First, the card has a common value element for dealers who plan on reselling the card; the common value element for non-dealers should be less predominant as some are likely to want the card for their collection not resale. Second, evidence from lab valuation suggests that as bidders gain experience with the market and the going market price, the variability of bids declines as many see the market price as an informative signal (e.g. Grether, 1994; Plott, 1996). Bids cluster as bidders learn about the market within which they trade. Therefore, we should reject the hypothesis that the bid variability of dealers is similar to the variability of non-dealers in the 1-good auction, $\text{Var}[b_{jone}^i] > \text{Var}[b_{dealer}^i] (j = H, A)$.

We consider three hypotheses to explore whether the context of choice affects that calibration functions for the Cal Ripken Jr. card, $b_{Aone}^i = f(b_{Hone}^i)$, $b_{Aten}^i = f(b_{Hten}^i)$, and $b_{Adealer}^i = f(b_{Hdealer}^i)$. First, for all three calibration functions, we test the notion that no bias exists in hypothetical behavior, that is, $\alpha=0$; $\beta=1.0$, where α represents the regression intercept; β represents the regression slope coefficient on hypothetical bid b_{Hone}^i or b_{Hten}^i ; the regressand is the actual bid, b_{Aone}^i or b_{Aten}^i . Second, the no-bias hypothesis implies that bids are symmetrical in that no extra bias exists when other goods or intensity of experience are present in the value elicitation process. We use a likelihood ratio test to determine if the coefficients generated from the 1-good treatment equal the coefficients from the 10-good treatment; if bias exists, we should reject the hypothesis that $\alpha_{one}=\alpha_{ten}$ and $\beta_{one}=\beta_{ten}$. Finally, we test whether experience affects the calibration function, We again use a likelihood ratio test to determine if the auction coefficients generated from non-dealers are the same as the coefficients from dealers; we should reject the hypothesis $\alpha_{one}=\alpha_{dealer}$ and $\beta_{one}=\beta_{dealer}$.

3. Results and discussion

Consider the general pattern of bidding behavior in the three treatments. Table 1 shows the mean and median bids and the demographic characteristics for participants in the 1-good, 10-good, and Dealer treatments. A one-way ANOVA test indicates that the respective samples for the three treatments do not differ by the socio-economic characteristics listed in Table 1, thereby assuring that bids across treatments differ due to treatment rather than demographic differences. The results show that the distribution of hypothetical bids lies to the right of the distribution of actual bids. A Wilcoxon matched-pairs signed ranks test rejects equality of the distributions at the 1 percent level for each

Table 1
Selected characteristics of auction participants

	1-Card auction (<i>N</i> =99)	10-Card auction (<i>N</i> =93)	Dealers (<i>N</i> =30)
Age ^a	34.2	31.4	33.9
Gender (% male) ^b	0.94	0.93	1.00
Education ^c	3.9	3.4	4.1
Income ^d	4.2	3.5	4.6
Experience ^e	8.9	9.4	9.8
Average hyp. bid (standard deviation)	142.02 (126.67)	91.71 (102.6)	208.80 (81)
Median bid	125.00	40.00	190.00
Average actual bid (standard deviation)	55.87 (82.9)	26.40 (52.2)	95.50 (88.1)
Median bid	5.00	0.00	92.50
Excluding 0s	75.00	35.00	140.00
Zero bids (%)	47 (47.4)	53 (56.9)	9 (30)

^a Age denotes actual age in years.

^b Gender denotes categorical variable (0–1): 0, if female, 1, if male.

^c Education denotes categorical variable (1–6): 1=grade 8 or less; 2=high school graduate; 3=2 years college; 4=other post-high school education; 5=4 year college; 6=graduate school.

^d Income categorical variable (1–8): 1=less than \$10,000; 2=\$10,000–\$19,999; 3= \$20,000–\$29,999; 4=\$30,000–\$39,999; 5=\$40,000–49,999; 6=\$50,000–\$47,999; 7=\$75,000–\$99,999; 8=\$100,000 or over.

^e Experience denotes actual years involved with sports cards.

treatment (1-good: $Z=-6.9$; 10-good: $Z=-7.7$; Dealer; $Z=-5.9$). Additionally, bias as revealed by mean central tendency shows the ratio of hypothetical-to-actual overbidding ranges from 2.2 to 3.5, depending on auction type. This degree of overbidding falls within the range of 1.0–10.0 observed in earlier work on hypothetical–actual bidding (see, Diamond and Hausman, 1994). The average overbidding exceeds that observed in Fox et al. (1998) irradiated and non-irradiated pork auctions, but is lower than in Neill et al.’s (1994) water color paintings and map auctions. These results reinforce the argument that people overstate their actual willingness to pay. Table 1 also shows that hypothetical bids from the 1-good auction required (0.39) less deflation than in the 10-good auction (0.28).

3.1. Market benchmark

Our auction market prices are reasonably close to the listed book value from Beckett. Table 2 presents the highest and second-highest hypothetical and actual bids for the Cal Ripken Jr. TT card in the three treatments and Beckett’s book value, \$350 (as of October 1995). For the Ripken 1982 TT card, hypothetical bids captured 143, 100, and 93 percent of the book value for the 1-good, 10-good and Dealer treatments while the actual bids garnered 97, 71, and 73 percent of the book value. Table 2 also lists the top two bids for the other 9-goods with their book values. Overall, the average percentage of book value captured by the auctions was over 162 percent for the hypothetical bids and about 94 percent for the actual bids.

Table 2
Top 2 hypothetical and actual bids

Treatment	Card	Hyp. bid (% of book)	Actual bid (% of book)	Book value
1-Card	Cal Ripken 1982 TT	(1) \$600 (2) \$500 (143)	\$350 \$340 (97)	\$350
Dealer	Cal Ripken 1982 TT	(1) \$330 (2) \$325 (93)	\$260 255 (73)	\$350
10-Card	Cal Ripken 1982 TT	(1) \$450 (2) \$350 (100)	\$300 \$250 (17)	\$350
	Cal Ripken 1982 Topps	(1) \$110 (2) \$110(138)	\$80 \$75 (94)	\$80
	Cal Ripken 1982 Donruss	(1) \$80 (2) \$60 (109)	\$50 \$45 (82)	\$55
	1982 TT Without Ripken	(1) \$70 (2) \$70 (108)	\$50 \$45 (96)	\$65
	Billy Ripken 1989 Fleer	(1) \$100 (2) \$90 (225)	\$50 \$35 (88)	\$40
	Troy Aikman 1989 Score	(1) \$110 (2) \$100 (222)	\$60 \$55 (122)	\$45
	Troy Aikman 1989 TT	(1) \$10 (2) \$7 (233)	\$5 \$3 (100)	\$3
	Troy Aikman 1989 Proset	(1) \$12 (2) \$8 (160)	\$7 \$4 (80)	\$5
	1989 TT without Aikman	(1) \$15 (2) \$13 (217)	\$8 \$5 (83)	\$6
	Michael Irvin 1989 Topps	(1) \$10 (2) \$8 (267)	\$6 \$5 (167)	\$3
Percent of book value (12 card mean)		167.9	93.84	

3.2. Internal consistency

Overall, the results suggest bidding behavior is broadly consistent with theoretical predictions. First, the inclusion of other goods reduced bids, both hypothetical and actual – mean and median hypothetical bids fell by 35 percent (\$142.02–\$91.71) and 68 percent (\$125–\$40), and mean and median actual bids fell by 53 percent (\$55.87–\$26.40) and 100 percent (\$75–\$0); excluding zero bids, the median fell by 53 percent (\$75–\$35). Using a Wilcoxon test, we reject the hypothesis that the populations for bids, hypothetical and actual, elicited in the 1-good auction were similar to bids elicited from the 10-good auction at the 5 percent level or better (hypothetical; $Z=-2.61$; actual; $Z=-2.06$). The observed decrease in values also falls within the range observed in previous studies on hypothetical values – 24–54 percent in Hoehn and Loomis (1993) and 60 percent in Cummings et al. (1994). Second, using a Moses test, we also reject the equal variance hypothesis at the 1 percent level for hypothetical and actual bids (hypothetical: $M=121$; actual; $M=96$). The dealer's knowledge of the common market value seemed to reduce the dispersion of their bids.

Table 3
Calibration functions

Variable	Model		10-Card auction		Dealer	
	1-Card auction					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Tobit	OLS	Tobit	OLS	Tobit
Constant	-8.3 (-0.7)	-20.0* (-1.9)	0.25 (0.1)	-0.19 (-0.8)	3.1 (0.3)	6.8 (0.1)
Hyp. bid	0.65** (4.7)	0.39** (5.1)	0.16 (1.8)	0.30** (5.8)	0.20 (1.9)	.27 (1.4)
Hyp. bid ²	-0.8E-0.3** (-2.5)	-	0.7E-0.3** (2.5)	-	0.8E-0.3** (-2.2)	-
Log likelihood	-559	-334	-448	-234	-210	-136

Dependent variable is actual bid.

^a *T*-statistics in parentheses.

** Significant at the 99 percent level.

* Significant at the 95 percent level.

3.3. Calibration functions

We use Tobit MLE and OLS to estimate the calibration functions⁶. Also, we include expansions in both square and square root of the hypothetical bid (hyp. bid) and present these estimates when significantly different than zero at the 5 percent level⁷. Estimates are corrected for heteroscedasticity in all auctions except the Dealer treatment.

Table 3 presents estimates of the calibration functions, both Tobit and OLS, $b^i_{Aone} = f(b^i_{Hone})$, $b^i_{Aten} = f(b^i_{Hten})$, and $b^i_{Adealer} = f(b^i_{dealer})$ across the three different auction. Models 1–2, 3–4, and 5–6 reflect the 1-good, 10-good, and Dealer treatments. The OLS estimates suggest the calibration functions are concave for both 1-card auction (dealer and non-dealer), suggesting that the relationship between hypothetical and actual reported valuations is an inverted-U shape. However, since these estimates may be biased and inconsistent, further investigation is warranted. In all Tobit specifications, the quadratic and square root terms are insignificant, suggesting that a linear calibration function is appropriate. Focusing on the Tobit estimates, we reject the no-bias hypothesis ($\alpha=0; \beta=1$) at the 1 percent level for all treatments. We also reject NOAA’s ‘divide by 2’ default hypothesis ($\alpha=0; \beta=0.5$) at the 1 percent level for the 1-good and 10-good treatments, and at the 10 percent level for the Dealer treatment.

This result suggests that calibration might have to proceed good-by-good, in which each good needs its own auction to capture the particular correlation between actual and

⁶ Since the dependent variable (actual bid) cluster at a finite lower limit (zero), the OLS distributional assumptions are violated and applying least squares to the data leads to biased and inconsistent estimates of the calibration function (Cragg, 1971). Since this is not a problem of censoring since bidders probably have a non-negative valuation for each good, one of two techniques can be used to estimate the calibration functions: Tobit maximum likelihood estimation (MLE) or Heckman’s two-step approach (i.e. Heckit).

⁷ We do not run a Box–Cox Tobit to test for non-linearities in the unknown calibration functions to avoid imposing an artificial monotonicity conditions.

hypothetical bidding. This raises a problem for calibration of non-deliverables. One potential solution is a two-step calibration process: collective hypothetical values are first calibrated to private hypothetical values for a closely related private proxy (e.g. adopt-a-species); and then calibrate the private hypothetical bids with private actual bids for the private proxy. For example, Harrison et al.'s (1997) wetland calibration project used a nature calendar, auctioned privately and collectively, to adjust for both free-riding and hypothetical bias. Many questions remain, however, about whether one calibration function can be transferred to a new context.

Using a likelihood ratio statistic, we reject the hypothesis that the coefficients from the 1-good treatment equal the 10-good treatment, $\alpha_{\text{one}} = \alpha_{\text{ten}}$ and $\beta_{\text{one}} = \beta_{\text{ten}}$, at the 1 percent level ($\chi^2 = 16.46$). A Mann–Whitney test also rejects pooling the data at the 5 percent level ($Z = -2.06$). Additionally, we reject the null hypothesis that the coefficient on the hypothetical bid is equal across the two auctions at the 12 percent level. To avoid constraining the variances to equality across auction types we use the dummy variable interaction approach within a heteroscedastic model. These results suggest the subjects bidding in the 1-good treatment have a greater tendency to overstate the hypothetical bid relative to bidding in the 10-good case.

To shed more light on the role of substitutes and complements in the 10-good treatment, we group the other good into three categories: (1) strong complements (Billy Ripken, TT set minus Cal Ripken); (2) strong substitutes (Cal Ripken Topps and Donruss); and (3) weak substitutes (five football cards); and estimate their calibration functions. All estimated slope coefficients from the Tobit model are significant at the 1 percent level. The slope coefficient for the two strong complements are 0.28 (Billy Ripken) and 0.16 (TT minus Cal Ripken); for strong substitutes, 0.24 (Cal Ripken Topps) and 0.23 (Cal Ripken Donruss); and for weak substitutes, 1.72 (Aikman Topps), 0.33 (Aikman TT), 1.38 (Aikman Proset), 1.11 (TT without Aikman), and 1.85 (Irvin Topps). A likelihood ratio test suggests these goods do not have the same calibration function at the 1 percent level ($\chi^2 = 9.14$). The parameter estimates suggest that the Cal Ripken 1982 TT affects weak substitutes differently than strong substitutes and complements. The calibration function is sensitive to the context of choice.

Table 4 shows that most bidders used a local or global bidding strategy, driving down bids for each individual good; where we define four strategies: global bidders – bids > 0 for 6 goods or more; local bidders – bids > 0 for 2–5 goods; solo bidders – bids > 0 for 1 good only; and no bidders – bids = 0 for all 10 goods. Table 4 also reveals, however, an inconsistent pattern of hypothetical and actual bidding strategies in the 10-good auction. Consistent bidding strategies are on the diagonal in Table 4; inconsistent strategies are off the diagonal. Nearly 55 percent (54 of 93) of the bidding strategies were inconsistent. Bidders systematically shifted from hypothetical global bidding to actual local, solo, and bidding strategies. Using the Wilcoxon signed ranks test, we reject the null hypothesis that the distributions of the hypothetical and actual bidding strategies are identical at the 1 percent level ($Z = -8.374$). Inconsistent bidding strategies imply many bidders in the hypothetical auction either violated the standard marginal efficiency conditions or viewed the binding budget constraint as dormant. If this assertion was false, bidders would not have reallocated their resources when their binding budget constraints were for real. A hypothetical bidding strategy that satisfied the standard efficiency and binding constraint

Table 4
Hypothetical and actual bidding strategies in the 10-card auction

		Actual bidding strategy				
		Global	Local	Solo	No	Σ
Hypothetical bidding strategy	Global	21	30	4	40	65
	Local	0	15	0	7	22
	Solo	0	0	0	0	0
	No	0	0	0	6	6
	Σ	21	45	4	23	93

Global: bid >0 for 6 cards or more; Local: bids >0 for 2–5 cards; Solo: bids >0 for 1 card only; No: bids=0 for all 10 cards.

conditions should not change in the actual auction but only 1 of 93 bidders was consistent.

We have mixed support for the null hypothesis that market experience does not affect the calibration function, $\alpha_{\text{one}} = \alpha_{\text{dealer}}$ and $\beta_{\text{one}} = \beta_{\text{dealer}}$. Using the likelihood ratio statistic, we cannot reject the null hypothesis at any reasonable level of significance ($\chi^2 = 1.5$). Results from the Mann–Whitney test, however, reveal that the null should be rejected at the 5 percent level ($Z = -2.77$). Also, we reject the hypothesis of identical slope coefficients across dealer/non-dealer auction types at the 5 percent level. More market experience implies, weakly, that the adjustment required is less stringent – knowledge dampens but does not eliminate the tendency to overstate hypothetical bids.

4. Concluding remarks

We design and implement a field experiment to calibrate in-sample hypothetical and actual values for an object with the desirable properties of being deliverable and familiar, that still has some qualities beyond the norm. While the results support the view the people overstate actual bids, the estimated calibration function to correct for this exaggeration is both good- and context-specific, that is, other goods and market experience matters. The calibration factors were 0.39 in the 1-good auction, 0.28 in the 10-goods auction, and 0.46 in the Dealer auction, all more stringent than the 0.6–0.9 range for irradiated meat products (Fox et al., 1998).

Good- and context-specific calibration suggests that one might be tempted to skip over the hypothetical question and go directly to the actual auction. But for policy debates over public goods, actual field auctions are constrained by one’s ability to deliver the goods. And while this might justify a common deflator such as NOAA’s divide-by-2 rule, our results show this could be off by 2–8 times – a sizable difference if aggregated to regional or national levels. Future research should explore whether (a) private goods can serve as reasonable proxies for public preferences, and (b) the burden of calibration can be reduced by clustering goods into a limited set of functions defined by the context of choice.

Acknowledgements

We thank the reviewers, Mike McKee, Bob Berrens, David Brookshire, Charlie Kolstad, Bob Deacon and seminar participants at New Mexico and University of California – Santa Barbara for their comments.

References

- Bishop, R., Heberlien, T., 1979. Measuring values of extramarket goods: Are indirect measures biased? *American Journal of Agricultural Economics* 61, 926–930.
- Blackburn, M., Harrison, G., Rutström, E.E., 1994. Statistical bias functions and informative hypothetical surveys. *American Journal of Agricultural Economics* 76, 1084–1088.
- Bohm, P., 1972. Estimating demand for public goods: An experiment. *European Economic Review* 3, 111–130.
- Bohm, P., 1984. Revealing demand for an actual public good. *Journal of Public Economics* 24, 131–151.
- Carson, R., Flores, N., Martin, K., Wright, J., 1996. Contingent valuation and revealed preference methodologies: Comparing the estimates for quasi-public goods. *Land Economics* 72, 80–99.
- Coursey, D., Hovis, J., Schulze, W., 1987. The disparity between willingness to accept and willingness to pay measures of value. *Quarterly Journal of Economics* 102, 679–690.
- Cragg, J., 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* 39, 829–843.
- Cummings, R., Brookshire, D., Schulze, W., 1986. Valuing environmental goods: An assessment of the contingent valuation method, Rowman and Allanheld, Totowa, NJ.
- Cummings, R., Ganderton, P., McGuckin, T., 1994. Substitution effects in contingent valuation estimates. *American Journal of Agricultural Economics* 76, 205–214.
- Cummings, R., Harrison, G., Rutström, E.E., 1995. Homegrown value and hypothetical surveys: Is the dichotomous choice approach incentive compatible? *American Economic Review* 85, 260–622.
- Diamond, P., Hausman, J., 1994. Contingent valuation: Is some number better than no number? *Journal of Economic Perspectives* 8, 45–64.
- Dickie, M., Fisher, A., Gerking, S., 1987. Market transactions and hypothetical demand data: A comparative study. *Journal of American Statistical Association* 82, 69–75.
- Fox, J., Shogren, J., Hayes, D., Kliebenstein, J., 1998. CVM-X: Calibrating contingent values with experimental auction markets. *American Journal of Agricultural Economics* (in press).
- Grether, D., 1994. Individual behavior and market performance. *American Journal of Agricultural Economics* 76, 1079–1083.
- Harrison, G., Beekman, R., Brown, L., Clements, L., Mc Daniel, T., Odom, S. and Williams, M., 1997. Environmental damage assessment with hypothetical surveys: The calibration approach, In: Bowman, M., Brannlund, R., Kriström, B. (Eds.), *Topics in Environmental Economics*, Kluwer, Amsterdam, forthcoming.
- Hoehn, J., Loomis, J., 1993. Substitution effects in the valuation of multiple environmental programs. *Journal of Environmental Economics and Management* 25, 56–75.
- Kagel's, J., 1995. Auctions: A survey of experimental research, In: Kagel, J., Roth, A. (Eds.), *Handbook of Experimental Economics*, Princeton University Press, Princeton, NJ, pp. 501–585.
- Krishna, V., Rosenthal, R., 1995. *Simultaneous Auction with Synergies*, Photocopy, Boston University.
- McAfee, P., McMillan, J., 1996. Analyzing the airwaves auction. *Journal of Economic Perspectives* 10, 159–176.
- McMillan, J., 1994. Selling spectrum rights. *Journal of Economic Perspectives* 8, 145–162.
- Melton, B., Huffman, W., Shogren, J., Fox, J., 1996. Consumer preferences for fresh food item with multiple quality attributes: Evidence from an experimental auction of pork chops. *American Journal of Agricultural Economics* 78, 916–923.
- Milgrom, P., Weber, R., 1982. A theory of auction and competitive bidding. *Econometrica* 50, 1089–1122.
- National Oceanic and Atmospheric Administration, 1994. Natural resource damage assessment: Proposed rules, *Federal Register*, 4 May, 59, pp. 23098–23111.

- National Oceanic and Atmospheric Administration, 1996, Natural resource damage assessments: Final rules, Federal Register, 5 January, 61, 439 pp.
- Neill, H., Cummings, R., Ganderton, P., Harrison, G., McGukin, T., 1994. Hypothetical surveys and real economic commitments. *Land Economics* 70, 145–154.
- Plott, C., 1996. Rational individual behavior in markets and social choice processes, In: Arrow, K. et al., Eds., *The Rational Foundations of Economic Behavior*, Macmillan (London) and St. Martin Press (New York).
- Randall, A., 1997. Calibration of CV responses: Discussion, In: Bjornstad, D., Kahn, J. (Eds.), *The Contingent Valuation of Environmental Resources*, Edgar Elgar, London, pp. 198–207.
- Seip, K., Strand, J., 1992. Willingness to pay for environmental goods in Norway: A contingent valuation study with real payment. *Environmental and Resource Economics* 2, 91–106.
- Shogren, J., 1990. The impact of self-protection and self-insurance on individual response to risk. *Journal of Risk and Uncertainty* 3, 191–204.
- Shogren, J., Shin, S., Hayes, D., Kliebenstein, J., 1994. Resolving differences in willingness to pay and willingness to accept. *American Economic Review* 84, 255–270.
- Vickrey, W., 1961. Counterspeculation, auctions and competitive sealed tenders. *Journal of Finance* 16, 8–37.