ELSEVIER
Economics Letters 60 (1998) 263–268
economics letters

Hypothetical–actual bid calibration of a multigood auction

John A. List[a], Michael Margolis[b], Jason F. Shogren[c,*]

[a] *Department of Economics, University of Central Florida, Orlando, FL 32765, USA*
[b] *Economics Department, Libbey Forum, Bates College, 44 Mountain Ave., Lewiston, ME 04240-6241, USA*
[c] *Department of Economics and Finance, University of Wyoming, Laramie, WY 82071-3985, USA*

Received 7 January 1997; received in revised form 24 April 1998; accepted 28 April 1998

Abstract

Evidence suggests the calibration of hypothetical and actual behavior is good-specific. We examine whether clustering commodities into mutual categories can reduce the burden. While we reject a common calibration across sets of commodities, a sport-specific calibration function cannot be rejected. © 1998 Elsevier Science S.A.

Keywords: Valuation; Calibration; Clusters

JEL classification: C9; Q0

1. Introduction

A nagging question in the nonmarket valuation literature is whether hypothetical bids send systematic but upwardly biased signals of underlying preferences for goods with intangible qualities (see Diamond and Hausman, 1994, and the cites therein). If so, researchers may be able to correct the bias of some goods by conducting nonhypothetical lab auctions, and then calibrating hypothetical and actual bidding behavior.[1] The results from some initial in-sample calibration experiments have been discouraging—evidence suggests the in-sample calibration function is commodity- and context-specific, implying that each hypothetical study would need its own lab auction to calibrate behavior (see, e.g., Fox et al., 1998 and List and Shogren, 1998).[2]

*Corresponding author. Tel.: +1-307-766-5430; Fax: +1-307-766-5090; E-mail: jramses@uwyo.edu

[1] The National Oceanic and Atmospheric Administration's proposed rules for natural resource damage assessment suggest a default calibration factor of two (NOAA, 1994). This deflator serves as an ad hoc placeholder until researchers develop an empirically based alternative. A key item on the research agenda is the nature of calibration for large-scale public goods. If calibration is infeasible because the researcher is unable to construct an experiment that would auction the good in the lab, then more information is needed on how calibration works for private goods and whether these functions are transferable to public goods.

[2] Diamond and Hausman (1994) note that the calibration factors implied by several studies vary from 1.5 to 10. Since these studies vary widely in method, they do not indicate that surveys administered according to a consistent protocol will fail to reveal consistent calibration. Fox et al. (1998), whose experimental results suggest that even with the same treatment, calibration factors may vary between commodities, poses a more serious caveat to calibration.

PII: S0165-1765(98)00117-7

The evidence has only been suggestive, however, since previous experiments have not been structured to test the hypothesis that goods share the same calibration function. This note presents such a test and explores whether clustering goods into common categories can reduce the burden of calibration. Using data from a field auction that elicited both hypothetical and actual bids for five baseball and five football sportscards, we reject the hypothesis that a common calibration function exists for all cards. This suggests that hypothetical bias varies significantly across commodities, even under identical experimental treatment. Our results, however, leave open the possibility that goods might be clustered on some objective basis. This finding provides some support for the supposition that each hypothetical study will not need a unique lab auction to calibrate bidding behavior.

2. Data

We use data from a field experiment at a 1995 Denver trading card show (List and Shogren, 1998). Sports trading cards are a useful object to explore calibration because they are familiar and deliverable, but still have some intangible qualities beyond a standard market good. Using a simultaneous sealed-bid, second-price auction (Vickrey, 1961; Melton et al., 1996), the field experiment followed a four-step experimental design: (1) inspection of the good(s), (2) hypothetical bid(s), (3) actual bid(s), and (4) debriefing. Ninety-three attendees followed this four step format and first submitted a set of hypothetical bids, and then a set of actual bids for 10 trading cards—five baseball and five football.[3] The hypothetical and actual auctions were identical except for the understanding that the actual bids were binding.[4] Table 1 presents the summary statistics for the hypothetical and actual bids for the 10 cards. Results from a Wilcoxon matched-pairs signed-ranks test rejects equality of the distributions of hypothetical and actual bids at the 1% level for each card (card 1: $z = -7.71$; card 2: $z = -7.76$; card 3: $z = -7.67$; card 4: $z = -7.44$; card 5: $z = -7.4$; card 6: $z = -6.9$; card 7: $z = -6.89$; card 8: $z = -6.48$; card 1: $z = -6.77$; card 10: $z = -6.31$). Hence, the distribution of hypothetical bids lies to the right of the distribution of actual bids, a pattern consistent with some earlier valuation exercises (e.g., Neill et al., 1994).[5]

3. Results and discussion

Given that the hypothetical–actual bid wedge persists, we consider whether a common calibration function exists for all cards, two clustered functions exist for baseball and football cards, or if a unique function is needed for each card. Since each subject submitted 10 bid pairs, one for each card, we estimated models allowing fixed or random effects to control for unobserved characteristics of

[3] Table 1-column 1 shows the contents of the bundles.

[4] See List and Shogren (1998) for a more complete discussion of the experimental design.

[5] As an aside, consider the default calibration scheme suggested in NOAA (1994), $\mu_A = \alpha_0 + \alpha_1 \mu_H$; where μ_A and μ_H are the mean actual and hypothetical bids, and the α_i are parameters which are constant across commodities. NOAA suggested that $\alpha_0 = 0.0$ and $\alpha_1 = 0.5$ until evidence proves otherwise. Our estimated equation is $\mu_A = -0.092_{(0.442)} + 0.291\mu_{H(0.12)} (N = 10)$, where the terms in parenthesis are standard errors. Thus we cannot reject either of the NOAA default values.

Table 1
Descriptive statistics[a]

Card	Actual bids			Hypothetical bids		
	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
Cal Ripken 1982 TT	26.40	52.207	0	91.71	102.668	40
Cal Ripken 1982 Topps	11.76	18.318	0	40.32	30.599	35
Cal Ripken 1982 Donruss	7.85	11.604	0	29.03	20.354	25
1982 TT w/o Cal Ripken	5.80	10.303	0	21.96	19.299	20
Billy Ripken 1982 Fleer	4.88	8.578	0	22.34	23.186	15
Troy Aikman 1989	0.82	1.170	0	3.32	2.662	4
Troy Aikman 1989 TT	10.85	16.674	0	29.81	32.489	20
Troy Aikman 1989 Proset	0.93	1.576	0	3.12	2.978	2
1989 TT w/o Aikman	1.20	1.9267	0	4.29	4.231	3
Michael Irvin 1989 Topps	0.90	1.399	0	2.87	2.736	3

[a]Sample includes actual and hypothetical bids from 93 participants. Each participant first submitted a hypothetical bid for each card and proceeded to place an actual bid for each card.

each subject. For brevity, we present results from only one of these models—the maximum likelihood estimator of the standard random effects.[6] Column 1 of Table 2 shows results from the equation:

$$A_{ij} = \sum_{j=1}^{10} \beta_j D_j + \sum_{j=1}^{10} \varphi_j H_{ij} D_j + \mu_i + \varepsilon_{ij}, \tag{1}$$

where A_{ij} and H_{ij} represent the actual and hypothetical bid of the i^{th} bidder for the j^{th} card, D_j is a dummy variable for card j, $\mu_i \sim N(0, \sigma_\mu)$ is the random disturbance characterizing the i^{th} bidder, and $\epsilon_{ij} \sim N(0, \sigma_\epsilon)$ is the bid-specific error term.

First, we test the restriction that all ten cards can be pooled into a single common calibration function (column 2 of Table 2):

$$A_{ij} = \alpha + \varphi H_{ij} + \mu_i + \varepsilon_{ij}. \tag{2}$$

We reject this restriction at the 1% level (LR tests: $\chi^2 = 29.13$; 18 d.f.). Rejection of the common

[6]A Hausman (1978) test failed to reject the orthogonality assumption ($\chi^2 = 3.46$; 19 d.f.), thus we do not report the less efficient fixed effects estimates. Nevertheless, results from the fixed effects model are qualitatively identical to those in Table 2. Note also that the bids are truncated. We do not present random effects Tobit estimates since the Tobit random effects likelihood function appears to be highly nonconcave with the present data. Furthermore, the Tobit procedure is inconsistent under virtually any relaxation of the underlying stochastic assumptions (Amemiya, 1985).

Table 2
Maximum likelihood random effects estimates[a,b]

Variable	(1) Ten groups	(2) General	(3) Sport-specific
Constant	−7.78	−1.27	−3.04
	(1.783)	(0.758)	(0.923)
D2	4.95		
	(2.635)		
D3	6.01		
	(2.727)		
D4	7.42		
	(2.510)		
D5	7.40		
	(2.392)		
D6	8.45		2.82
	(2.610)		(0.928)
D7	7.86		
	(2.375)		
D8	8.51		
	(2.463)		
D9	8.27		
	(2.44)		
D10	8.35		
	(2.466)		
Hypo	0.37	0.34	0.35
	(0.013)	(0.009)	(0.010)
HypoD2	−0.01		
	(0.043)		
HypoD3	−0.04		
	(0.064)		
HypoD4	−0.09		
	(0.067)		
HypoD5	−0.14		
	(0.056)		
HypoD6	−0.33		0.01
	(0.483)		(0.033)
HypoD7	−0.01		
	(0.041)		
HypoD8	−0.31		
	(0.433)		
HypoD9	−0.21		
	(0.305)		
HypoD10	−0.26		
	(0.470)		
LL	−3672.74	−3687.31	−3681.05

[a]Dependent variable is actual bid.[b]Standard errors in parentheses.

calibration function null implies that hypothetical bias is not systematically identical across baseball and football cards. This is potentially damaging to the idea that calibration functions can be readily and costlessly transferred across goods (Blackburn et al., 1994). If a common calibration function eludes football and baseball cards—relatively similar goods—how confident can one be that a common function would exist across a broader spectrum of goods?

Second, we consider whether clustering the cards into good-specific groups can reduce the burden on calibration. Instead of ten calibration functions, we test the restriction that the ten cards can be pooled into two sport-specific clusters, baseball and football cards, such that two calibration functions exist (column 3 of Table 2):

$$A_{ij} = \beta\sum_{j=1}^{5} D_j + \phi\sum_{j=6}^{10} D_j + \varphi\sum_{j=1}^{5} H_{ij}D_j + \rho\sum_{j=6}^{10} H_{ij}D_j + \mu_i + \varepsilon_{ij}. \tag{3}$$

This looser restriction cannot be rejected at any conventional significance level ($\chi^2 = 16.62$; 16 d.f.). Rejection of the pooling null (Eq. (2)), but nonrejection of the quasi-pooling null (Eq. (3)), could be due to many reasons, but one intuitive explanation is that many collectors trade baseball cards more intensely and view baseball cards as more central to their collection.[7] Therefore, many collectors may view baseball and football cards as distinct commodities. In any case, the evidence so far leaves open the possibility that similar commodities have similar calibration functions.

Overall, the evidence suggests that some measurable correlation exists between intentions and actual behavior. This measurable correlation appears to be commodity- and context-specific. Although one size does not necessarily fit all, a potentially profitable research direction is to identify goods with tangible parameters that can be clustered into common calibration functions. Perhaps these calibration clusters can replace the need for actual auctions, but we are a long way from that day. Future in-sample calibration experiments involving goods further out on the continuum of intangibility may get us closer to a systematic understanding of why and to what degree hypothetical bids diverge from real commitments.

4. Concluding remarks

When the tanker *Exxon Valdez* ran aground on Bligh Reef in March 1989, it reinvigorated the debate about the usefulness of hypothetical valuation exercises. The persistent fear that hypothetical bids send systematically upwardly biased signals of real preferences triggered a search for an *ex post* calibration function to deflate these bids. Using data from a field experiment, we find that an interchangeable calibration function did not exist for a set of sports trading cards, but that one common calibration function might exist if the cards are clustered into their respective sports. The ability to cluster goods provides some optimism that each survey might not need its own lab auction to calibrate hypothetical intentions with real economic commitments. More research exploring the robustness of our findings would be most useful.

[7]According to representatives from *Scoreboard*™ Inc., *Fleer/Skybox*™ Inc., and *Pinnacle*™ Inc., baseball cards make up an estimated 65–80 percent of the sportscard market, suggesting collectors are more familiar with baseball cards or view baseball cards as more central to their collection or both.

References

Amemiya, T. 1985. Advanced Econometrics. Harvard University Press, Cambridge.

Blackburn, M., Harrison, G., Rutström, E.E., 1994. Statistical bias functions and informative hypothetical surveys. American Journal of Agricultural Economics 76, 1084–1088.

Diamond, P., Hausman, J., 1994. Contingent valuation: is some number better than no number?. Journal of Economic Perspectives 8, 45–64.

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.

Hausman, J., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271.

List, J., Shogren, J., 1998. Calibration of the difference between actual and hypothetical valuations in a field experiment. Journal of Economic Behavior and Organization, in press.

Melton, B., Huffman, W., Shogren, J., Fox, J., 1996. Consumer preferences for fresh food items with multiple quality attributes: evidence from an experimental auction of pork chops. American Journal of Agricultural Economics 78, 916–923.

National Oceanic and Atmospheric Administration. Natural Resource Damage Assessments; Proposed Rules. Federal Register, 59(5) (1994) 1062–1191.

Neill, H., Cummings, R., Ganderton, P., Harrison, G., McGuckin, T., 1994. Hypothetical surveys and real economic commitments. Land Economics 70, 145–154.

Vickrey, W., 1961. Counterspeculation, auctions and competitive sealed tenders. Journal of Finance 16, 8–37.