Valuation on the Frontier: Calibrating Actual and Hypothetical Statements of Value

Richard A. Hofler and John A. List

The lack of robust evidence showing that hypothetical behavior directly maps into real actions remains a major concern for proponents of stated preference nonmarket valuation techniques. This article explores a new statistical approach to link actual and hypothetical statements. Using willingness-to-pay field data on individual bids from sealed-bid auctions for a $350 baseball card, our results are quite promising. Estimating a stochastic frontier regression model that makes use of data that any contingent valuation survey would obtain, we derive a bid function that is not statistically different from the bid function obtained from subjects in an actual auction. If other data can be calibrated similarly, this method holds significant promise since an appropriate calibration scheme, *ex ante* or *ex post*, can be invaluable to the policy maker that desires more accurate estimates of use and nonuse values for nonmarket goods and services.

Key words: calibration, contingent valuation, stochastic frontier.

Benefit-cost analysis remains the dominant paradigm used throughout the public sector. Yet, a recurring issue in properly estimating the total benefits of nonmarket goods and services is whether hypothetical statements map into real behavior. Some recently published studies provide evidence that suggests important differences exist between responses from real and hypothetical valuation questions.[1] This observation triggered a search for an *ex ante* and *ex post* procedure to correct the systematic bias between intentions and actions in valuation exercises. Recent technology using *ex ante* procedures has produced some strong evidence that hypothetical bias can be overcome by using a "cheap talk" scheme (see, e.g., Cummings and Taylor; List). *Ex post* calibration has also shown some signs of promise, as work due to Blackburn, Harrison, and Rutström, Fox et al., and Mansfield has provided some hope that hypothetical bias can be rectified using appropriate *ex post* statistical techniques.[2]

At this early stage, we believe that both *ex ante* and *ex post* techniques show considerable promise, and we hope that researchers pursue both avenues as potential solutions to hypothetical bias. In this study, we extend *ex post* calibration schemes by calibrating hypothetical statements to real values using a relatively new statistical approach—stochastic frontier estimation. An important advantage of the frontier approach is that only data that are typically gathered in a contingent valuation survey, such as stated values and other demographic characteristics, are needed to carry out the calibration exercise. Our method is therefore usable by anyone who has access to contingent valuation data and desires to estimate

Richard Hofler is professor, Department of Economics, University of Central Florida and John List is professor, Department of Agricultural and Resource Economics, University of Maryland.

[1] See, e.g., Neill et al., Frykblom, and List and Shogren (2001). But, some summary studies suggest that behavior across hypothetical and actual regimes is closely associated (Carson et al.). The interested reader should also see the debate between Cummings et al. (1997) and Haab, Huang, and Whitehead and Smith.

[2] Elaborated more fully in Harrison, the use of estimated bias functions to value public goods is at least as old as Kurz. Nevertheless, perhaps panelist Roy Radner at the NOAA public meeting reinvigorated the notion of using *ex post* calibration techniques to estimate the demand for nonmarket goods. For example, Radner asked one speaker: "what would be a practical method, if any, of taking the results of a CVM willingness to pay and adjusting them . . . in order to come to a damage assessment? How would one go about that?" (NOAA, 1992, p. 99). Later Radner inquired: "are there things that one can do when one does the CVM, if one were to do it, and that would minimize this bias and, secondly, enable one to estimate it?" (p. 100).

use and/or nonuse values.[3] The initial empirical results are promising. Using real bidding data as a benchmark, our statistical model provides individual bid functions (derived from hypothetical responses) that are *not statistically different* from the benchmark bid functions obtained from subjects participating in an actual auction. If other data can be calibrated similarly, then this method holds significant promise since an appropriate calibration scheme, *ex ante* or *ex post*, can be invaluable to the policy maker that desires more accurate estimates of use and nonuse values for nonmarket goods and services.

The remainder of our article is organized as follows. The next section describes the data and our field experiment. The following section outlines our estimation methodology. Then there is a section on empirical results and finally a concluding section.

Data

We use data from a field experiment at a sportscard show in Denver, CO in December 1995 (List and Shogren, 1998) to compare behavior across hypothetical and actual Vickrey sealed-bid second price auctions, with particular attention paid to the statistical link between bids across the two regimes. As has been argued in previous studies (e.g., List), unlike traditional field studies, field experiments represent an exciting opportunity to test the validity and relevance of the predictions of economic theory because one can take advantage of the exogenous variation in the variables of interest, allowing for relatively novel tests of economic theory. While field experiments may not be as "clean" as laboratory experiments, they have the virtue of resembling natural economic phenomena as closely as possible. In addition, they are able to check the robustness of laboratory results in a natural setting, where the mathematical assumptions of the theory cannot necessarily be guaranteed to hold. Hence, controlled field studies provide a useful middle ground between the sterile environment of the laboratory and the unruly nature of uncontrolled field data.

We conducted the Vickrey second price auctions with "ordinary" consumers—nonsportscard dealers—each bidding against rivals of

the same type. The auctioned sportscard was a Cal Ripken, Jr. 1982 *Topps Traded* PSA-graded 9 rookie baseball card, which had a book value of $350. As elaborated on in List and Shogren (1998), the field experiment can be described in three simple steps: (*a*) inspection of the good, (*b*) hypothetical bid, and (*c*) actual bid. In Step 1, the monitor briefly explained that in the first stage of the experiment we were hypothetically auctioning off the Ripken baseball card displayed on the dealer table. The participant could pick up and visually examine the card. The monitor worked one-on-one with the participant and no time limit was imposed on his or her inspection of the card. After the participant had examined the card, Step 2 began. In Step 2, the participant read an instruction sheet that consisted of two parts: (*a*) a short socioeconomic survey (obtaining information on subject age, education, income, and gender), and (*b*) a bidding sheet. The participant was asked to submit a hypothetical bid stating the maximum that she was willing to pay for the card. Instructions for the bidding sheet stated that the exchange mechanism was a sealed-bid second price auction. After the participant filled out the survey and placed the bidding sheet in an opaque box, she stepped to a second table 15 feet away for a follow-up auction.

In Step 3, monitor B informed the participant that she now had the chance to actually bid on the Ripken card that she had just examined in Step 2. Monitor B gave the participant a second bidding sheet for the real auction. After the monitor answered all questions about the auction, the participant placed her sealed bid into a second opaque box. To guarantee that we did not get a second hypothetical bid, we asked each participant to acknowledge their actual bid with a signature and valid telephone number where they could be contacted.[4] Care was taken to avoid contamination of the results by any ordering effects (e.g., sealed boxes, monitors not handling or observing bids). Cummings, Harrison, and Rutström (1995) found no evidence of an ordering effect in a dichotomous choice setting.

---

[3] As elucidated in Mansfield, many other calibration approaches require actual market data from weakly complementary goods or the identification of a proxy good for the nonmarket good in question.

[4] One valid concern is whether the participant actually believes she will have to pay if she wins the auction. In this particular auction, we received the payment within three days of notifying the winner. More generally, we have run approximately seven field studies in the sportscard market and this sort of problem has not emerged. One explanation for this result is that we are dealing with actual subjects that commonly participate in these sorts of auctions and a "culture" of honoring bids has developed.

Table 1. Characteristics of Auction Participants

	Mean (Std. Dev.)
Income	4.20
	(2.01)
Age	34.20
	(11.84)
Gender (% male)	0.94
	(0.24)
Education	3.90
	(1.51)
Experience	8.90
n	99

Notes:

1. Income denotes categorical variable (1–8): (1) Less than $10,000, (2) $10,000 to $19,999, (3) $20,000 to $29,999, (4) $30,000 to $39,999, (5) $40,000 to $49,999, (6) $50,000 to $74,999, (7) $75,000 to $99,999 (8) $100,000 or over.
2. Age denotes actual age in years.
3. Gender denotes categorical variable: 0 if female, 1 if male.
4. Education denotes categorical variable (1) Eighth grade or less, (2) High School (3) 2-Year College, (4) Other Post High School, (5) 4-Year College, (6) Graduate School Education.
5. Experience denotes number of years involved with the sportscard market.

Table 1 provides a breakdown of selected characteristics of the auction participants. Sample sizes and central tendencies of the variables are similar in nature to data in other sportscard field studies, such as List. An important attribute of our field experiment is that in addition to gathering important demographic data, our within-person design permits us to obtain both hypothetical and real bids from each subject. In this sense, when calibrating hypothetical bids, we have an internal benchmark, real bids, with which to compare our estimated bid function.

A Stochastic Frontier Approach

We employ the stochastic frontier methodology developed by Aigner to examine the baseball card data. The stochastic frontier was developed, and continues to be largely used, in a production efficiency context. Since Hofler and Polachek, however, the scope of stochastic frontier applications has expanded beyond production. Examples of nonproduction applications include estimating reservation wages, the natural rate of unemployment, and the labor market effort.

Given that the sportscard data are obtained from a sealed-bid auction and are cross-sectional in nature, we specify a stochastic frontier

$$(1) \qquad Y_i^H = \mathbf{X}_i \beta + v_i + u_i, \qquad i = 1, \ldots, n$$

which represents a "bid function" for each person. Y_i^H represents the hypothetical (observed) bid for person i; $\mathbf{X}_i$ is a row vector of person-specific bid-determining characteristics; and β is a column vector of regression coefficients. The error term v is assumed to be normally distributed with $E(v_i) = 0$, and $\text{VAR}(v_i) = \sigma_v^2$. The one-sided error term is u.

In considering the distributional assumption for the one-sided error term u an interesting choice emerges. One puzzle that has drawn much attention in the valuation literature is whether hypothetical values are equivalent to real (actual) values. If one assumes that the commodity is well defined but unfamiliar, the model outlined in Crocker and Shogren provides a theoretical structure for a systematic overstatement of willingness to pay (WTP). This implies a one-sided random error term u that takes on only nonnegative values, suggesting hypothetical values greater than or equal to actual values. The inequality reverses if one follows the theory of Hoehn and Randall, which suggests hypothetical bids are systematically less-than-true values when the good is not clearly defined and the subject has little time to develop a value.

This lack of theoretical consensus creates a serious dilemma for the researcher interested in calibration. Not only must an accurate method to calibrate hypothetical statements be found, but the direction to calibrate must also be determined. Inherent in our approach is a test for the correct direction of misstatement (if any) in hypothetical values. The researcher applies the test of the misstatement direction *while* estimating the stochastic frontier model. Results of the statistical test indicate whether respondents are systematically over- or understating their actual values.

The foundation for this directional test is described as follows: the preferred stochastic frontier estimation method is to calculate ordinary least squares (OLS) estimates as starting points for the maximum likelihood (ML) estimates. Before beginning ML estimation, the skewness of the OLS residuals is checked. Waldman shows that the ML estimator for the stochastic frontier model's one-sided error variance (σ_u^2) equals zero if the OLS residuals are skewed in the opposite direction from what is assumed. This implies that OLS is a MLE, and the MLE step is unnecessary. In such cases, the researcher has made the wrong assumption concerning the direction of the one-sided error

and must specify a stochastic frontier model with the opposite error specification.[5]

For calibration of our data, if we assume that hypothetical values overstate actual values and the OLS residuals are negatively skewed (they should be positively skewed if the assumption is correct), then the ML estimate of σ_u^2 equals zero. Conversely, if we assume that hypothetical statements are lower bound estimates of real values and the OLS residuals are positively skewed (they should be negatively skewed if the assumption is correct), then the same result emerges. In both cases, the researcher's response is to estimate the stochastic frontier model with the opposite error specification.[6]

We applied the Waldman skewness test to our Ripken auction data. The correct specification was confirmed to be in accord with the theory of Crocker and Shogren. Hence, we model u as a one-sided random error term that takes on only nonnegative values and represents the distance by which hypothetical bids exceed actual bids.[7] As previously mentioned, while some summary studies suggest that behavior across hypothetical and actual regimes is closely associated (Carson et al.), our finding is consistent with some sets of experimental data that suggest hypothetical bids are greater than actual bids (List and Shogren, 2001).

We therefore consider actual (true) bids as "frontier bids" in our model. That is, they are unobserved but are potentially estimable. The model to be estimated, equation (1), can therefore be seen as

$$(2) \qquad Y_i^H = Y_i^A + u_i, \qquad i = 1, \ldots, n$$

where Y_i^H is an observed hypothetical bid; Y_i^A is the actual (frontier) bid that is not observed; $u_i \geq 0$ with $E(u_i) = \mu$, and $\text{VAR}(u_i) = \sigma_u^2$ and

$$(3) \qquad Y_i^A = \mathbf{X}_i \beta + v_i \qquad i = 1, \ldots, n.$$

Based upon our specification of $u_i \geq 0$, our model implies that

$$(4) \qquad Y_i^H \geq Y_i^A, \qquad i = 1, \ldots, n$$

meaning that actual bids are less than or equal to hypothetical bids.

The error term u_i is a measure of the gap between true and hypothetical bids for the ith individual: as u_i approaches zero the gap decreases, or hypothetical values → real values. An estimate of $u_i = 0$ implies that the person has truthfully revealed her true bid when asked a contingent valuation question. As u_i increases, the disparity between the hypothetical and actual bid function is greater.

There are four customary choices for the distribution of the one-sided error component u: (a) u is i.i.d. $N(0, \sigma_u^2)$ and truncated at zero (a "half-normal" distribution); (b) u is exponential; (c) u is i.i.d. $N(\mu, \sigma_u^2)$ where $\mu \neq 0$ and truncated at zero (a "truncated normal" distribution) and u is gamma. A fair question to ask at this point is, "Will the estimates of the gap between true and hypothetical bids be sensitive to which of the four error specifications is chosen?" It is likely that the gap estimates will be sensitive to error choice. However, some investigators have found that the choice matters little.[8] Notwithstanding what they have concluded, we chose to explore this issue by estimating models with *each* of the four error term specifications. The parameters of the model are estimated via ML with the start values derived from initial OLS estimates.

Finally, note that Jondrow et al. show that an estimate of u_i can be expressed as the expected value of u_i given ε_i. Extending this notion, we see that the equivalent of their formula for the truncated normal u case is:

$$(5) \qquad E(u_i \mid \varepsilon_i) = \frac{\sigma \Gamma}{(1 + \Gamma^2)} \left[\frac{f(\mu^*)}{F(\mu^*)} - \frac{\varepsilon_i \Gamma}{\sigma} \right]$$

where $\mu^* = \frac{\varepsilon_i \Gamma}{\sigma} + \frac{\mu}{\sigma \Gamma}$, $\Gamma = \frac{\sigma_u}{\sigma_v}$, $\sigma^2 = \sigma_u^2 + \sigma_v^2$, and f and F are the standard normal density and distribution functions, respectively.[9] Since Γ, σ_u^2, and σ_v^2 are parameters estimated by maximum likelihood and ε_i can be computed for each observation, person-specific u_i estimates

---

[5] More fully, if this condition emerges, one of two possibilities exists: either (*a*) the model is not well specified or (*b*) the data are inconsistent with the model. Once the researcher is confident that the model is well specified, the only remaining conclusion is (*b*).

[6] The astute reader will question whether it is possible for the alternative specification to *also* yield the ML estimate of σ_u^2 equaling zero. In the authors' experience with numerous data sets, this has never been an outcome.

[7] As a test, an attempt was made to estimate the stochastic frontier model with the opposite error specification. The expected result of the ML estimate of σ_u^2 equaling zero occurred. This provides evidence that this simple skewness test could possibly be used to determine the direction of the misstatement of hypothetical values in other samples.

[8] Greene (1990) estimated four sets of efficiency values with means between 0.8766 and 0.9011. Ritter and Simar concluded that the choice among the densities is largely immaterial. Finally, see Greene's (2000) Table 9.2.

[9] Estimates of u_i will be calculated by different versions of (6) for each of the error specifications.

can be obtained from equation (6). With estimates of u_i in hand, the error term v_i can be calculated since ε_i can be obtained easily. Next, each person's calibrated true bid can be calculated. This permits the researcher to adjust hypothetical bids closer to the true (unobserved) bid for each person if necessary.

Empirical Results

To illustrate the potential value of the stochastic bid frontier in calibrating hypothetical bids, we use the baseball card data of List and Shogren (1998) discussed above. Here, we push the data somewhat harder and estimate an individual's actual value from the relationship between stated hypothetical values and a vector of regressors. Regressors in vector X are in the spirit of Mansfield and Blackburn et al., and include measures of gender, age, education, income, and sportscard market experience for each of the ninety-nine subjects (see Table 1 for descriptive statistics). Our estimation procedure follows three steps: (*a*) estimate the stochastic bid frontier using individual hypothetical bids as the dependent variable (DV) and the five demographic variables as independent variables, (*b*) obtain an empirical estimate of the bid gap ratio (true bid as a percentage of hypothetical bid) for each individual, and (*c*) calculate the calibrated true bid using the estimated bid gap ratio and hypothetical statement.

At this stage, a reader might fairly ask whether different forms of the bid frontier yield different calibrated true bids. As this seems quite possible, we explore the estimation of two different forms. Specifically, we estimate a log-linear bid frontier ($\ln y_i = \Sigma x_i \beta_i +$ error term) and a translog frontier.[10] The net effect of our model specification decisions is that we estimate eight models: four different error specifications in a log-linear model and four error specifications in a translog model.

Concerning Step 3, we are aware of two methods that make intuitive sense. One method (method A) involves adjusting each person's hypothetical bid by his or her own estimated bid gap ratio. For instance, if the hypothetical bid for person *j* is $50 and her estimated gap ratio is 0.12 (the true bid is estimated to be 12% of the hypothetical bid), then

the calibrated true bid is $6. If the hypothetical bid for person *k* is $600 and his estimated gap ratio is 0.90 (the true bid is estimated to be 90% of the hypothetical bid), then the calibrated true bid is $540.

An alternative method (method B) entails adjusting every hypothetical bid by the overall mean gap ratio. For instance, if the hypothetical bid for person *j* is $50 and the mean estimated gap ratio is 0.33 (on average across the sample true bids are estimated to be 33% of the hypothetical bids), then the calibrated true bid is $17. If the hypothetical bid for person *k* is $600 and the mean estimated gap ratio is 0.33, then the estimated true bid is $199. This method is similar in spirit to the *initial* National Oceanic and Atmospheric Administration's (NOAA) blue-ribbon panel recommendation that hypothetical bids be deflated using a "divide by 2" rule unless bids can be calibrated using actual market data (NOAA, 1994, 1996). One important advantage of both A and B approaches is that the data demands for estimating the actual bid function are small—only data that are usually collected in a typical contingent valuation survey are necessary. No additional data are needed to implement the stochastic frontier model.

At this stage, we have described three decisions affecting the estimation of calibrated bids: four error specifications, two functional forms, and two calibration methods. The consequence of implementing all of these combinations is that we generate sixteen distinct calibrated bid series. This strategy allows us to explore the robustness of our bid frontier method. And, as previously mentioned, a novel characteristic of our field data is that we have information on both hypothetical and actual bids for each individual. This allows us to compare the calibrated bids with the actual bids. Hence, we can examine whether the adjusted hypothetical bids match the bids actually made in the Vickrey second-price auction. This internal validity test serves as a means of investigating the usefulness of our stochastic bid frontier approach.

Approaches A and B Both Achieve Statistical Significance

A very reasonable concern of our approach is whether the true bid distribution is skewed. If so, then some/all of what we measure as hypothetical bias would be an overstatement of that bias. Note that we are assuming the errors

---

[10] We also considered estimating a Fourier flexible form, but this requires estimating seventy-one coefficients in our case. With ninety-nine observations, the degrees of freedom are too small to permit doing so.

are the sum of a symmetric and nonsymmetric distribution. The first assumption is, of course, what most researchers typically assume and is, in fact, a reasonable assumption in most cases. In other words, we are not assuming that the bids are symmetrically distributed. Rather, we are assuming that the impact of factors such as random events is symmetric.

Yet, it is certainly possible that even after controlling for causal variables, the distribution of WTP values could remain skewed. In Waldman et al., the authors observe that for inexperienced bidders in an experimental situation, bids from contingent valuation surveys are likely to be skewed. To rectify this nuance, they recommend using a Box-Cox transformation. They also report that, using real data from three studies, they find the estimated Box-Cox parameter to be quite small, suggesting that a log-bid specification is appropriate.

Following the guidance of Waldman et al., we conducted two Box-Cox tests.[11] First, we applied the Box-Cox test to the model with *hypothetical* bids as the dependent variable. The estimated value of lambda is not significantly different from zero, suggesting that the log-hypothetical bid specification is appropriate. This result is consonant with the findings in Waldman et al. Second, we tested the model with *actual* bids as the dependent variable. The estimated value of lambda is significantly different from zero and close to one, suggesting that the actual bids are not skewed.

Digging deeper into the issue of bid skewness, we chose to also estimate a frontier model using the *actual* bids as the dependent variable. One of the likelihood function's parameters is gamma, which is the ratio of the variance of the one-sided error over the variance of the composed error: $\gamma = \sigma_u^2/(\sigma_u^2 + \sigma_v^2)$.[12] Gamma = 0 suggests that no inefficiency (skewness) exists because the numerator variance must be zero in that case. Estimating a model using actual bids, the null that gamma = 0 cannot be rejected. This suggests that the actual bids are symmetrically distributed conditional on the observed explanatory variables. This, in turn, suggests that the asymmetry we observe in the stochastic frontier model with hypothetical bids is due to what we claim—participants' hypothetical bids systematically overstate their

actual bids. Accordingly, we estimate a log-WTP frontier model.

As stated above, we estimate sixteen different bid frontiers. In order to conserve space, we do not present all sixteen estimated models.[13] Instead, we discuss the results that are generally consistent across all sixteen models. Of the five independent variables (age, education, gender, income, and experience), gender (positive coefficients) and experience (negative coefficients) almost always significantly affect the nature of the hypothetical bid function at conventional significance levels. The estimated positive coefficients on gender suggest that males offer larger hypothetical bids than do females. Concerning years of experience in the sportscard market, the negative coefficients imply that more experienced bidders tend to place lower hypothetical bids than those consumers that have little sportscard market experience.

From the sixteen estimated models, we calculate sixteen series of calibrated true bids. Table 2 displays the means of each of the series and the mean hypothetical and actual bids. The table also presents results of *z*-tests of the null hypothesis that the population mean bid from the actual auction (or hypothetical auction) equals the population average for each calibrated series.[14] Each Method A calibrated bid is calculated using individual bid gap ratios as the adjustment factor. Each Method B bid is computed with the series mean estimated bid gap ratio as the adjustment factor.

To begin the discussion of these results, we note that hypothetical bids have a mean of $142 and actual bids have a mean of about $56. These first moments suggest that hypothetical bids observed in the Vickrey second price auction were about 2.5 times greater than the actual bids from the same auction institution. This result is consistent with some previous studies that compare hypothetical and actual behavior—on average subjects overstate their preferences by a factor of about 3 in hypothetical willingness-to-pay settings (List and Shogren, 2001). Table 2 reveals a number of characteristics of the calibrated bids. First, the average of all calibrated bids is just under $68. These bids are an average 1.2 times the size of the actual bids observed in the real

---

[11] Thanks to an astute referee for suggesting these tests. All of these tests used the truncated normal specification for the one-sided error. We have no reason to believe that the outcomes would differ with other specifications.

[12] Note that this gamma differs from the gamma in (6), which is used to estimate u_i.

[13] All results not displayed in this article are available from the authors upon request.

[14] These *z*-tests (N is large) use sample variance for the actual and hypothetical bids and, following Horrace and Schmidt, the sampling error for the calibrated bids (because they are estimated).

Table 2. Bids: Hypothetical, Actual, and Calibrated (Calibrated bids estimated 16 different ways: by model, error specification, and method)

Model	Error Specification	Method	Mean Calibrated Bid	Distance Above/Below Mean Actual Bid	Mean Calibrated Bid Different from Mean Actual Bid?*	Mean Calibrated Bid Different From Mean Hypothetical Bid?*
Log-linear	Half-normal	A	75.90	20.02	No	Yes
Log-linear	Exponential	A	75.99	20.12	No	Yes
Log-linear	Truncated normal	A	78.99	23.11	No	Yes
Log-linear	Gamma	A	106.34	50.47	Yes	Yes
Log-linear	Half-normal	B	45.33	−10.55	No	Yes
Log-linear	Exponential	B	45.41	−10.47	No	Yes
Log-linear	Truncated normal	B	47.18	−8.69	No	Yes
Log-linear	Gamma	B	68.57	12.70	No	Yes
Translog	Half-normal	A	84.34	28.46	No	Yes
Translog	Exponential	A	80.72	24.84	No	Yes
Translog	Truncated normal	A	88.46	32.58	Yes	Yes
Translog	Gamma	A	71.45	15.58	No	Yes
Translog	Half-normal	B	55.33	−0.54	No	Yes
Translog	Exponential	B	54.54	−1.33	No	Yes
Translog	Truncated normal	B	59.51	3.64	No	Yes
Translog	Gamma	B	47.41	−8.46	No	Yes
Overall mean			67.84	11.97		
Mean actual bid			55.87			
Mean hypothetical bid			142.02	86.15		

*Testing H_0: Population mean for actual/hypothetical bids = population mean for calibrated bids at 5% level.
These *z*-tests (N is large) use the sample variance for the actual and hypothetical bids and, following Horrace and Schmidt, the sampling error for the calibrated bids because they are estimated.

auction: $68 compared to $56. One can readily see that from a central tendency perspective, the calibrated mean estimate is much closer to the true bids than are the hypothetical bids obtained from the contingent auction (recall that the hypothetical bids are at an average 2.5 times the size of the actual bids observed in the real auction: $142 compared to $56.)

We chose to estimate two different functional forms in order to examine the robustness of our results. Average bids from both the log-linear model and the translog model are approximately $68. Thus, there appears to be little advantage in either model if one looks merely at the relationship between the actual bids and the mean calibrated bids. There also appears to be no systematic advantage in one form over the other in over-(under) calibration. Five of the calibrated series from each model overestimate the actual bids and three of the series underestimate them, on average.

Bisecting the sixteen calibrated bid series by either method exposes tendencies in each method to over- (under) calibrate. All eight of the series calculated by method A provide means that *exceed* the average actual bids: the mean for all method A bids is nearly $83, which is approximately 1.5 times greater than the mean actual bid. Conversely, six of the eight method B series *under* estimate the actual bids: the mean for those calibrated bids is about $53, which is 95% of the mean actual bid. From this perspective, method B provides a more conservative, and more accurate, calibrated value than does method A.

Parsing by both method and function reveals a potentially important lesson. The mixture of method and function that yields the most accurate calibrated bids is the method B-translog combination. The series that is closest to the actual bids (on average) is the translog half-normal model using method B, which is an astounding 54 *cents* away from the actual bids. Furthermore, the four series from the method B-translog combination are 97% of the actual bids, on average. This is a remarkable performance from a sample containing only five

independent variables, all of which should be easily obtainable from most contingent valuation surveys.

Most importantly, both calibrated bid series outperform the hypothetical series in terms of estimating actual bids. Using z-tests at the conventional $p < 0.05$ level, we fail to reject the null hypothesis that there is no difference between the mean actual bid and each of the mean calibrated bids (A and B) in fourteen of the sixteen cases. Alternatively, we can reject the mean equality of the hypothetical and actual bids (and hypothetical and all sixteen series of calibrated bids) from Methods A and B at the $p < 0.01$ level. These results suggest that, in general, the calibrated bid function via both Methods A and B *does not* differ from the average actual bid observed in the Vickrey second price auction, but does significantly differ from the hypothetical bid function. This result holds promise in that the policy maker interested in making informed decisions might need a calibration technique that can adequately adjust hypothetical responses obtained from a contingent survey.

Concluding Comments

In this study we use field data from a sportscard auction to link hypothetical and actual bidding behavior for a private good. Our raw data, which are consistent with several past studies, support the view that people overstate actual values when asked a contingent valuation question. The contribution of the article, therefore, is to derive statistical techniques that meaningfully adjust these hypothetical bids to better represent actual bids. Using a calibration function derived from a stochastic frontier regression model, we find promising results. Our calibrated individual bid functions closely correspond to actual bid functions obtained in a real auction.

We hope that our empirical methodology, which only requires data found in most contingent valuation surveys, will represent the beginning of a research agenda to discover an appropriate calibration method to estimate true values. Future research should explore whether (*a*) similar results can be produced using other data, (*b*) the burden of calibration can be reduced by clustering goods, and (*c*) calibration functions are transferable across individuals and goods. Researchers have made use of some of these issues, but given the high stakes associated with enhancing the credibility of nonmarket valuation techniques, further work is imperative.

References

Aigner, D.J., C.A.K. Lovell, and P. Schmidt. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6(1977):21–37.

Blackburn, M., G. Harrison, and E.E. Rutström. "Statistical Bias Functions and Informative Hypothetical Surveys." *American Journal of Agricultural Economics* 76(1994):1084–88.

Carson, R.T., N.E. Flores, K.M. Martin, and J.L. Wright. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." *Land Economics* 72(1996):80–99.

Crocker, T.D., and J.F. Shogren. "Preference Learning and Contingent Valuation Methods." In F. Dietz, F. Van der Ploeg, and J. Van der Straaten, eds. *Environmental Policy and the Economy*. Amsterdam: North-Holland Publishers, 1991, pp. 85–107.

Cummings, R., S. Elliott, G. Harrison, and J. Murphy. "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy* 105(1997):609–21.

Cummings, R., G. Harrison, and E.E. Rutström. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?" *American Economic Review* 85(1995):260–66.

Cummings, R., and L. Taylor. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review* 89(1999):649–65.

Fox, J., J.F. Shogren, D. Hayes, and J. Kliebenstein. "CVM-X: Calibrating Contingent Values with Experimental Auction Markets." *American Journal of Agricultural Economics* 80(1998):455–65.

Frykblom, P. "Hypothetical Question Modes and Real Willingness to Pay." *Journal of Environmental Economics and Management* 34(1997):275–87.

Greene, W.H. "A Gamma-Distributed Stochastic Frontier Model." *Journal of Econometrics* 46(1990):141–64.

——. *Econometric Analysis*, 4th ed., Upper Saddle River, NJ: Prentice-Hall, 2000.

Haab, T.C., J.C Huang, and J.C. Whitehead. "Are Hypothetical Referenda Incentive Compatible? A Comment." *Journal of Political Economy* 107(1999):186–96.

Harrison, G. "Experimental Economics and Contingent Valuation." Economics Working Paper 96-10, Division of Research, The Darla Moore School of Business, University of South Carolina, October, 1996.

Hoehn, J., and A. Randall. "A Satisfactory Benefit Cost Indicator from Contingent Valuation." *Journal of Environmental Economics and Management* 14(1987):226–47.

Hofler, R.A., and S.W. Polachek. "Ignorance in the Labor Market: A New Approach for Measuring Information Content." Proceedings of the 1982 Business and Economic Statistics Section, American Statistical Association, pp. 422–25.

Horrace, W.C., and P. Schmidt. "Confidence Statements for Efficiency Estimates from Stochastic Frontier Models." *Journal of Productivity Analysis* 7(1996):257–82.

Jondrow, J., C.A.K. Lovell, I. Materov, and P. Schmidt. "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model." *Journal of Econometrics* (1982):233–38.

Kurz, M. "Experimental Approach to the Determination of the Demand for Public Goods." *Journal of Public Economics* 3(1974):329–48.

List, J.A. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards." *American Economic Review* 91(2001):1498–1507.

List, J.A., and J.F. Shogren. "Calibration of the Difference between Actual and Hypothetical Valuations in a Field Experiment." *Journal of Economic Behavior and Organization* 37(1998):193–205.

——. "Calibration of Willingness-to-Accept." *Journal of Environmental Economics and Management* 43(2001):219–33.

Mansfield, C. "A Consistent Method for Calibrating Contingent Value Survey Data." *Southern Economic Journal* 64(1998):665–81.

National Oceanic and Atmospheric Administration (NOAA). *Contingent Valuation Panel, Public Meeting.* Wednesday, August 12, 1992, Certified Official Transcript Department of Commerce, 1992, 283 pp. plus attachments.

——. *Natural Resource Damage Assessment: Proposed Rules.* Federal Register, 59(4 May 1994): 23098–111.

——. *Natural Resource Damage Assessments: Final Rules.* Federal Register, 61(5 January 1996): 439.

Neill, H., R. Cummings, P. Ganderton, G. Harrison, and T. McGuckin. "Hypothetical Surveys and Real Economic Commitments." *Land Economics* 70(1994):145–54.

Ritter, C., and L. Simar. "Another Look at the American Electrical Utility Data." Working Paper, CORE and Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, 1997.

Smith, V.K. "Of Birds and Books: More on Hypothetical Referenda—Comment." *Journal of Political Economy* 107(1999):197–200.

Waldman, D. "A Stationary Point for the Stochastic Frontier Likelihood." *Journal of Econometrics* 18(1984):275–79.

Waldman, D, G.W. Schulze, G. McClellan, and J. Lazo. "Sources of Error in Contingent Valuation of Environmental Commodities." In D. Bjornstad and J. Kahn, eds. *Contingent Valuation of Environmental Commodities.* Cheltenham, UK: Edward Elgar Publishing, Ltd., 1996, pp. 67–95.