Sawtooth Software

RESEARCH PAPER SERIES

Estimating Willingness to Pay (WTP) Given Competition in Conjoint Analysis

Bryan Orme
Sawtooth Software, Inc.

3210 N. Canyon Rd., Provo, Utah
+1 801 477 4700
www.sawtoothsoftware.com

ESTIMATING WILLINGNESS TO PAY (WTP) GIVEN COMPETITION IN CONJOINT ANALYSIS

BRYAN ORME
SAWTOOTH SOFTWARE

This article describes how to simulate Willingness to Pay (WTP) in a more realistic and focused way than either the common algebraic approach or the two-product simulation approach. The common approaches don't consider competition and tend to overstate WTP. We recommend simulating product enhancements against a competitive set of alternatives (including the possibility of a None alternative) together with bootstrap sampling for estimation of confidence intervals around WTP. We introduce a generalizable and powerful extension called Sampling Of Scenarios (SOS) for estimating WTP that can be tailored to make certain detailed assumptions regarding the firm's product as well as competitive reactions in the marketplace. These new features are implemented in Sawtooth Software's desktop market simulator available within Lighthouse Studio and also as standalone Choice Simulator software.

INTRODUCTION AND MOTIVATION

Since the inception of conjoint analysis, researchers and their clients have sought intuitive ways to quantify the preference for attribute levels in monetary terms (e.g., Willingness to Pay or WTP). We should note that in economic studies, "paying" could involve other currencies such as time or travel distance and the approach we describe could be used to estimate WTP on other such currencies.

Historically, rather than reporting WTP, we at Sawtooth Software have preferred to quantify the impact of attribute levels on choice as changes in share of preference via sensitivity simulations. However, we also are frequently asked to deliver WTP. Over the last decade, these requests have only increased. Many consultants and other software packages compute WTP, but depending on the approach used, the monetary amounts are often overstated. The common approaches to WTP tend to overstate it, since they do not explicitly consider competition or the ability to opt out (choose the None). They also tend to average across respondents rather than focusing WTP more relevantly on respondents on the cusp of choosing the enhanced product features.

To promote better practice and more reasonable WTP results, we recommend procedures outlined below considering competition for estimating WTP via market simulation. These procedures are now available within our conjoint analysis market simulation software package, though researchers who have programming capabilities could implement them in other widely-used tools. Even though we've created easy-to-use software features to compute WTP, we caution researchers regarding pitfalls involved in interpreting WTP results, many of which we discuss below.

Failure to Account for Competition

In our opinion, failure to account for competitive alternatives is the main weakness in most commonly implemented WTP approaches and can lead to exaggerated WTP. In a 2001 paper later incorporated within the book, *Getting Started with Conjoint Analysis*, (Orme 2001, Orme 2004) we gave an example based upon the 1960s TV show, *Gilligan's Island,* illustrating how failure to account for competition can inflate WTP estimates. The cast is marooned on the island and a boat with capacity for two passengers appears on the scene ready to sell passage back to civilization to the highest bidders. The rich Mr. Howell appears willing to pay millions of dollars for passage for his wife "Lovey" and himself—until a second equally seaworthy boat appears offering the ride home for $5000. Mr. Howell of course chooses the $5000 option. The point of this illustration is that even though Mr. Howell is willing and able to pay over a million dollars, the availability and price of substitute goods in the marketplace means the firm (the boat) cannot capture this amount. If WTP is meant to represent the amount buyers are willing to pay the firm for enhanced features *in the current marketplace*, its calculation should account for competition including the None alternative (if available).

Two common approaches to WTP estimation do not consider competition or the ability to opt out and often can lead to inflated estimates of WTP:

1. The ***algebraic approach*** computes dollars per utile from the price function (the price utilities) and uses this to convert differences in utility between other attribute levels to monetary equivalents. This may be done at the individual level using HB utilities and the WTP estimates for the sample can be made more robust by taking medians (rather than means) across respondents.
2. The ***two-product market simulation approach*** simulates respondents choosing between just two versions of the product: one with and one without the enhanced feature. The price for the enhanced version of the product is adjusted upward via trial and error (or using our simulator's =SOLVEFORSHARE function) until the shares are distributed 50/50. The price difference that equalizes the shares of preference is taken as WTP. (Note: if the first-choice rule for simulating choice is used, the algebraic approach with medians and the two-product market simulation approach lead to identical WTP estimates. However, we generally recommend the logit share of preference approach.)

Simulation-Based WTP with Proper Competitive Context

Twenty years ago, Rich Johnson and this author recommended a *competitive set market simulation approach* for estimating WTP that accounts for competitive alternatives in the marketplace as well as the None alternative (Orme 2001). The approach involves first simulating[1] market choice for the unenhanced version of the

[1] For WTP estimation via simulations, we generally recommend using the Share of Preference (Logit equation) approach for estimating likelihood of choice for competing alternatives at the individual level, then averaging those results across respondents. First Choice, Randomized First Choice, and First Choice on the Draws could be used, but we would tend to prefer the Share of Preference approach operating on individual-level logit-scaled utilities for WTP estimation.

firm's product compared to a realistic set of competitive offerings and the None alternative. Next, we enhance the firm's product with a new feature and via trial and error (or using our simulator's SOLVEFORSHARE function). The increase in price that drives the share of preference for the firm back to the original base case share prior to feature enhancement is taken as the WTP. In real marketplaces, buyers are rarely limited to just one brand to obtain enhanced product features; they can select from many alternatives to achieve the same or compensating product benefits. Or, they can opt out. When competition is accounted for, WTP estimates are more realistic than methods that ignore competition.

HB-MNL and other MNL methods scale the utilities to be optimal for making predictions within the same context as the questionnaire's choice questions. Thus, if four alternatives were shown per question, the simulation predictions are more accurate when making predictions among a richer set of four alternatives rather than two alternatives as in the two-product market simulation approach.

We do not claim to be the first to propose using the market simulator with a competitive set of alternatives to find WTP as described in the previous paragraph. Rich Johnson, founder of Sawtooth Software, recommended it to me in the mid-1990s. Recently, I consulted via email with three very experienced conjoint researchers who were active in conjoint analysis in the 1980s (David Lyon, Keith Chrzan, and Joel Huber) and they agreed with Rich that estimating WTP using competitive simulation scenarios as I've described just seemed natural to do, given market simulation capabilities. They cannot recall being inspired regarding this by any specific article that detailed or originated the approach, as it just seemed common sense.

Upstream Steps to Improving WTP Analysis

Although not the subject of this article, I should note that hypothetical bias (e.g., respondents not spending real money in questionnaires or having to live with the consequences of their choices), interviewing the wrong people, and poor questionnaire design can inflate WTP estimates. We've outlined some ideas for improvement in these respects in our book, *Becoming an Expert in Conjoint Analysis* (Chrzan and Orme 2017). Noisy/bad data also can lead to exaggerated WTP and steps should be taken to remove respondents who appear to be answering randomly or completely ignoring price (Allenby et al. 2014). In extreme cases, we've found that data cleaning for speeders/random responders can remove as much as 50% of the data, though in our experience it's more typical to need to clean 15% to 25% of the sample.

Utility Constraints on Price

For most product categories that we'd expect to follow the law of supply and demand, we recommend constraining price to have negative slope (e.g., monotonically decreasing utilities as price increases) in estimation prior to computing WTP. Without price utility constraints, respondents with reversed price utilities could seem to choose a product with even higher likelihood as we increase the price. In some cases without price constraints, neither increasing nor decreasing the price of an enhanced product can drive its share back to the original share of preference prior to the product enhancement.

In such unusual cases, WTP via the competitive simulation approach doesn't yield a solution.

Proper Interpretation of WTP

As with other methods, our approach to WTP expresses monetary differences relative to a reference (base) level of an attribute. For example, if we've included three levels of speed (low, medium, and high) we consider a reference level (such as low speed) and estimate the value of the other two levels with respect to the reference level. For example, the relative WTP estimates may be:

Low speed	N/A (reference level)
Medium speed	$50 (relative to Low speed)
High speed	$120 (relative to Low speed)

Non-Additivity of WTP

Most approaches for estimating WTP for attribute levels focus on a single change in a product feature rather than a series of simultaneous feature improvements involving multiple attributes. A common error in interpreting WTP is to assume that WTP is additive across independent attributes. For example, if each of six features has an estimated WTP of $50 when *individually* and independently enhancing a base case product, it would be extrapolating beyond the assumptions of our WTP approach to conclude that the WTP for all six features *simultaneously* added to a base case product is 6 x $50 or $300.

Simply summing WTP values fails to account for diminishing marginal returns for cumulative product improvements (which is accommodated by the sigmoidal shape of the logit function). Summing WTP values also fails to consider increasing resistance due to buyers' budgetary constraints where increasing the price by cumulative amounts may very well push the utility function for price into a new region of the utility function that reflects greater price sensitivity. We can account for this by doing WTP analysis for multiple features taken simultaneously; it just requires simulating the firm's base case product followed by a version of the product enhanced by *multiple* features (again vis-à-vis relevant competition and the None alternative) and finding the indifference price via trial and error or using Sawtooth Software's automated SOLVEFORESHARE function.

New Advances in Simulating WTP for Conjoint Analysis

Next, we'll describe the main advances within Sawtooth Software's simulator for estimating WTP.

1. We've automated a search procedure for finding the change in price that leads to share of preference indifference (i.e., equality in preference shares) between enhanced and base case versions of the firm's product when those are placed in competition with other alternatives and the None.

2. Previously, there wasn't a straightforward way using our software to compute confidence intervals when using the recommended market simulation approach for computing WTP. We've implemented bootstrap sampling[2] to achieve confidence interval estimates of WTP.
3. For instances where the competitive set isn't well known or if the researcher wants to account for uncertainty in competitive reactions, we propose and introduce repeated Sampling Of Scenarios (SOS) with varying feature characteristics and pricing for both the firm's product as well as competitors.

CONTRASTS TO OTHER WTP APPROACHES

The two common approaches to estimating WTP described earlier in this article (the *algebraic approach and* the *two-product simulation approach*) are unrealistic, failing to account for one or more of the following:

- The firm usually doesn't hold a monopoly on enhanced features; competitors can also provide the enhanced features, sometimes at a price lower than the firm hopes to charge.
- Competitors can provide other combinations of features or prices that may attract buyer preference despite the feature enhancements made by the firm.
- Competition doesn't necessarily remain static but may react to the firm's product enhancements in myriad ways, either rational or irrational.
- Consumers aren't forced to buy anything, but usually can opt out (choose the None alternative).
- WTP primarily should consider the respondents likely to switch to or away from the firm's product (those on the cusp of buying).
- WTP should depend on the firm's current positioning and price. A base case price of $300 could lead to a different WTP for an enhanced feature than a base case price of $400.

Some approaches to economic valuation of product features focus on estimating additional profits to the firm due to product enhancement, accounting for competitive reaction and assuming a game theory framework where the firm along with competitors are guided by profit maximization goals until achieving Nash Equilibrium. Such an approach usually requires knowing or assuming feature costs for each of the players (such that profit may be computed), which is nearly impossible to ascertain in most

[2] We've long advocated using HB estimation for conjoint models and for ease of implementation and efficiency for practitioners in the trenches we've relied upon summary point estimates rather than the granular draws to predict respondent choices. We admit this is a simplification that departs from a full Bayesian treatment for representing uncertainty and confidence bounds. Instead, we employ bootstrap sampling to estimate confidence intervals for the simulation-based WTP method. Another simplification is to conduct bootstrap sampling within the same HB estimation run. Formally, bootstrap sampling should be done with new HB estimation runs (a separate HB estimation for each bootstrap sample), rather than just bootstrap sampling the posteriors within the same HB estimation run. We compare results for those two approaches in Appendix B.

situations. Neither of the two competitive simulation approaches to WTP we describe below require knowledge of costs.

The Fixed Competitors Simulation Approach

When the competitive set is known or may be approximated, the researcher can specify the firm's base case product and price along with the fixed characteristics of competitors within a simulation scenario. An exhaustive set of competitors doesn't need to be specified; but the important players in the market should be represented. As is best practice for conjoint market simulations, we recommend using high-quality individual-level utilities from hierarchical Bayesian (HB) estimation, Mixed Logit, or similar estimation methods. To estimate WTP for features associated with the firm's base case product, we employ the preference share (indifference) approach which finds the change in price associated with a product enhancement that drives the share of preference for the firm back to its original preference prior to making the enhancement. When an attribute is not the focus of WTP estimation, we return it to its base case specification for the firm's product. We can repeat the simulation multiple times using bootstrap sampling to estimate confidence intervals for WTP of the features. We recommend at least 300 bootstrap samples for reasonable estimates of confidence intervals and 1000 or more bootstrap samples if high precision is desired.

The Sampling Of Scenarios (SOS) Approach

The Sampling Of Scenarios (SOS) approach is a useful generalized approach when there isn't certainty as to the base case product specifications or when there is uncertainty about competitor composition and reactions. What makes the SOS approach different from the fixed competitors approach is that we repeatedly sample among randomly selected competitive positioning as well as random variations in the firm's product for which we are estimating WTP. For each sampled scenario, we estimate WTP in the same way we've described above: through finding the equalization price for the enhanced product that sets its share back to the original share prior to enhancement.

Our approach to SOS[3] can involve more than just a generalized random selection of features. If specified, it can account for certain assumptions (constraints) regarding the firm's focus offering or the competitors' features and pricing. For example, we can assume the firm cannot change its brand or (optionally) the level of one or more other attributes. We can use a researcher-defined base price for the firm's offering, or we can generalize it to cover the entire price range. If the competitive offerings cannot include certain attribute levels, we can specify those exclusions (e.g., the competitors cannot assume the brand level associated with the firm). If the firm believes it can hold a monopoly or patent on certain other feature(s), then the competitors can be prohibited from taking on such feature(s).

[3] Dave Lyon, the reviewer for this paper, suggested (and we agree) that the SOS approach could be used for general sensitivity analysis for attribute levels. This is where we specify competitive scenarios and observe how changes to the base case levels of the client's product affect the share of preference outcomes.

The reader may wonder how long it takes the software we've developed to perform WTP estimation for all attributes in a study for, say, 1000 Sampling Of Scenario draws. It takes around 5 to 45 minutes for the typical commercial CBC data sets we've experimented with. If performing bootstrap sampling with SOS draws, it can take 9 times as long as this if using the software's defaults, meaning a run that potentially takes a long lunch break to overnight to finish.

The SOS method seems robust to variation in the number of competitors used in the simulation scenario. (Appendix A)

Example Results for Conjoint Data Sets

We have tested WTP estimation using the new Sampling Of Scenarios (SOS) approach for nine conjoint data sets. All nine cases used HB estimation and we constrained price to be negative. For comparison, we also report the two common approaches for computing WTP that don't assume any competition (algebraic approach with medians[4] and two-product[5] simulation).

The results are fairly consistent across CBC datasets. The SOS approach tends to lead to the lowest estimates of WTP of the three methods we tried and also allows us to go beyond limitations of the other two approaches. For illustration, one of the datasets led to the following average WTP values:

	Algebraic Approach to WTP (medians)	2-Product Simulation Approach to WTP	SOS Approach vs. 5 competitors
TV Data Set (n=382):	$106	$99	$79

WTP for this TV Data Set is fairly representative of the results across the nine datasets, where the SOS approach usually leads to the lowest estimates of WTP and the algebraic approach usually leads to the highest WTP estimates.

To add more color to the analysis and demonstrate the additional capabilities of the SOS approach to WTP, we can estimate WTP assuming an exclusive patent on a feature. Let's imagine our product has a patent on Channel Blockout technology. This means that we can specify in the software that the random draws of competitors cannot take on this level. In that case, WTP for Blockout technology increases from $56 in the general case where competitors can include this feature to $87 if we hold an exclusive patent.

[4] For the algebraic approach with medians, we generally used part-worth price coding and constrained price to be negative. We calculated the dollars per utile by referencing the lowest and highest price levels.

[5] For the 2-product simulation approach, we generally simulated the two products starting at a price point about in the middle or lower third of the price range. We felt this would better approximate the average price sensitivity across the price function than, for example, always starting the simulations referencing the lowest price.

WTP for Blockout with and without Patent

With Patent:	$87
Without Patent:	$56

As a second example of the strength of the SOS approach to WTP, we can isolate WTP for Blockout technology holding a level of a different attribute constant, such as brand. There are three brands in this old CBC study collected in the mid-1990s: JVC, RCA, and Sony. From past experience with this data set (via Latent Class analysis), we know that respondents who prefer Sony tend to be less price sensitive than those preferring the other brands. We can run WTP analysis using the SOS approach, where the WTP product always carries (in turn) the Sony, RCA, or JVC brands and the random draws of competitors take on the other two brands. The resulting WTP for Blockout by brand is:

WTP for Blockout by Brand

JVC	$54
RCA	$50
Sony	$66

Below, we report WTP results for all nine data sets, where we've indexed each WTP estimation to 1.0 for comparison.

TV Data Set (n=382):

	Algebraic Approach to WTP (medians)	2-Product Simulation Approach to WTP	SOS Approach vs. 5 competitors
Mono to Stereo	1.12	1.11	0.77
No Blockout to Blockout	1.09	0.99	0.92
No Picture-in-Picture to PIP	1.14	0.99	0.87
Column averages:	1.12	1.03	0.85

Cessna Airplane Data Set (n=539):

	Algebraic Approach to WTP (Medians)	2-Product Simulation Approach to WTP	SOS Approach vs. 5 competitors
Brand1 to Brand4	1.18	1.10	0.71
A2L3 to A2L2	0.95	0.93	1.13
A3L1 to A3L2	1.05	1.03	0.92
A3L1 to A3L3	1.03	1.02	0.95
Column averages:	1.05	1.02	0.93

Phone Data Set (n=586):

	Algebraic Approach to WTP (Medians)	2-Product Simulation Approach to WTP	SOS Approach vs. 5 competitors
Brand5 to Brand4	0.80	0.95	1.24
A3L1 to A3L2	1.07	1.07	0.86
A4L1 to A4L3	1.01	1.08	0.92
A5L4 to A5L1	0.87	1.04	1.09
A5L4 to A5L3	0.95	1.05	1.09
A6L4 to A6L1	1.03	1.05	0.92
Column averages:	0.95	1.04	1.01

INDIAA Data Set (n=1202):

	Algebraic Approach to WTP (Medians)	2-Product Simulation Approach to WTP	SOS Approach vs. 5 competitors
Brand9 to Brand4	0.97	0.81	1.22
Brand9 to Brand5	1.23	0.78	0.99
Brand9 to Brand6	0.96	0.64	1.39
A2L6 to A2L4	1.23	1.12	0.65
A2L6 to A2L8	1.82	0.87	0.32
Column averages:	1.24	0.84	0.91

Cruise Line Data Set (n=600):

	Algebraic Approach to WTP (Medians)	2-Product Simulation Approach to WTP	SOS Approach vs. 5 competitors
Carnival to Norw	0.83	1.29	0.88
Inside to Ocean	1.17	1.14	0.69
Older to Newer	1.02	1.08	0.90
11 days to 7 days	0.87	1.13	1.00
Column averages:	0.97	1.16	0.87

Cons Data Set (n=120):

	Algebraic Approach to WTP (Medians)	2-Product Simulation Approach to WTP	SOS Approach vs. 5 competitors
A1L1 to A1L2	0.75	1.10	1.15
A3L6 to A3L4	1.41	0.87	0.71
A3L6 to A3L5	1.20	1.05	0.74
Column averages:	1.12	1.01	0.87

Chspr Data Set (n=356):

	Algebraic Approach to WTP (Medians)	2-Product Simulation Approach to WTP	SOS Approach vs. 5 competitors
Brand4 to Brand3	1.60	0.81	0.58
A2L2 to A2L3	0.96	0.99	1.05
A3L1 to A3L2	1.30	1.13	0.57
A4L2 to A4L1	1.79	0.79	0.42
A5L1 to A5L2	1.31	0.98	0.70
Column averages:	1.39	0.94	0.66

Flat Screen TV Data Set (n=951):

	Algebraic Approach to WTP (Medians)	2-Product Simulation Approach to WTP	SOS Approach vs. 5 competitors
Vizio to Samsung	0.75	0.96	1.29
1080p to 4K	0.62	1.07	1.31
No HDR to HDR	0.91	0.98	1.11
60Hz to 120Hz	0.81	1.43	0.77
3HDMI to 4HDMI	0.86	1.25	0.90
Column averages:	0.79	1.14	1.08

Study1 Data Set (n=420):

	Algebraic Approach to WTP (Medians)	2-Product Simulation Approach to WTP	SOS Approach vs. 5 competitors
Brand3 to Brand1	1.32	1.05	0.63
Brand3 to Brand2	1.16	1.09	0.75
A2L3 to A2L1	1.04	1.05	0.92
A3L1 to A3L2	1.31	1.01	0.69
Column averages:	1.09	1.03	0.88

Across data sets, the Algebraic approach tends to lead to higher WTP estimates. Counting how many times each approach led to the *highest* estimated WTP, we find:

Algebraic Approach (Medians)	6 out of 9
2-Product Simulation Approach	3 out of 9
Sampling Of Scenarios (SOS) Approach	0 out of 9

Averaging across the indices for the nine studies, the summary relative WTP prices are:

Algebraic Approach (Medians)	1.09
2-Product Simulation Approach	1.03
Sampling Of Scenarios (SOS) Approach	0.88

Referring to the relative WTP price indices, we see that the SOS approach leads to WTP values that are on average 14% lower than the 2-product simulation approach and 20% lower than the Algebraic approach with medians.

Bootstrap Sampling for Confidence Intervals

We can develop confidence intervals via bootstrap sampling. Recall that we are using HB estimation, so we have individual-level utilities. Bootstrap sampling involves sampling with replacement (repeatedly) as many respondents as are in the original data set. Note that sampling with replacement means that some of the original respondents will not appear in a given bootstrap sample, and others will appear two or more times. As an example, the INDIAA data set has 1202 respondents. We can create hundreds of samples of that data set each involving 1202 respondents via bootstrap sampling. The standard deviation of the WTP values across those resamples provides an unbiased estimate of the standard error. Then, the mean +/- 1.96 times the standard error defines the 95% confidence interval range.

Here are confidence interval results using bootstrap sampling for the INDIAA data set:

INDIAA Data Set (n=1202):

	SOS Approach vs. 5 competitors	Standard Error	Lower 95% Confidence Interval	Upper 95% Confidence Interval
Brand9 to Brand4	$45.0	$7.2	$30.8	$59.1
Brand9 to Brand5	$33.8	$3.4	$27.1	$40.4
Brand9 to Brand6	$10.6	$3.7	$3.4	$17.9
A2L6 to A2L4	$43.2	$4.2	$35.0	$51.5
A2L6 to A2L8	$4.5	$2.3	$0.1	$8.9

Conclusions

Willingness to Pay (WTP) has been challenging for the research community to get right. Although conjoint analysis gives us the right kind of data from choices within realistic-looking market scenario contexts, the common approaches to estimating WTP from conjoint analysis data have weaknesses. Those common approaches include the algebraic method and the 50/50 two-product simulation approach. Estimating WTP using conjoint market simulators that incorporate a realistic set of relevant and appropriate competition (including the None option) leads to more realistic and lower estimates of WTP. The market simulation approach to WTP considering competition focuses the estimation on respondents who are on the cusp of choice, rather than averaging results across respondents who may have relatively little interest in a given feature enhancement. The Sampling Of Scenarios (SOS) extension to the competitive

simulation approach allows us to either generalize WTP considering all possible competitive reactions, or to incorporate specific assumptions involving exclusivity of brand name or feature enhancements (e.g., due to a patent). Our results across nine commercial CBC datasets show that the market simulations approach considering a rich set of competitors obtains WTP estimates on average 20% lower than the common algebraic approach.

AREAS FOR FUTURE RESEARCH

Individual-level utilities from HB estimation are derived from a mixture of individual-level and upper-level (aggregate) information. Thus, we recognize that bootstrap sampling among lower-level HB utilities from a single HB estimation run may understate the true sampling variability in the WTP estimates, because we haven't re-estimated the utilities using HB for each resample. Rather, we are just taking a resampling among the HB utilities that have been estimated just once leveraging the full respondent sample. Estimating HB for each bootstrap sample would add a very large amount of time (typically 3 to 10 minutes for each resample). For conjoint studies that have a healthy number of choice tasks relative to parameters to estimate, the individual-level utilities rely less on the upper-level information, so confidence intervals may not be understated by much. However, for sparse conjoint datasets (relatively few choice tasks per individual relative to parameters to estimate), the upper-level model can have a fairly large influence on the individual-level estimates. We compare bootstrapped WTP results for confidence intervals between our approach and a more complete treatment involving re-estimation of the HB utilities for each resampling in Appendix B. But, more work could be done across a variety of datasets.

Running HB with useful external covariates can enhance the heterogeneity across respondent utilities. Thus, using covariates may provide more accurate (and wider) WTP confidence interval estimates when using our bootstrap approach than using plain vanilla HB estimation without covariates. The more sparse the conjoint data, the more helpful covariates could be in reflecting appropriate heterogeneity. We hypothesize that with a healthy number of choice tasks per respondent (15 or more choice tasks) and a few useful covariates related to preference, confidence interval estimates may not change much whether estimating HB separately for each bootstrap sample or not.

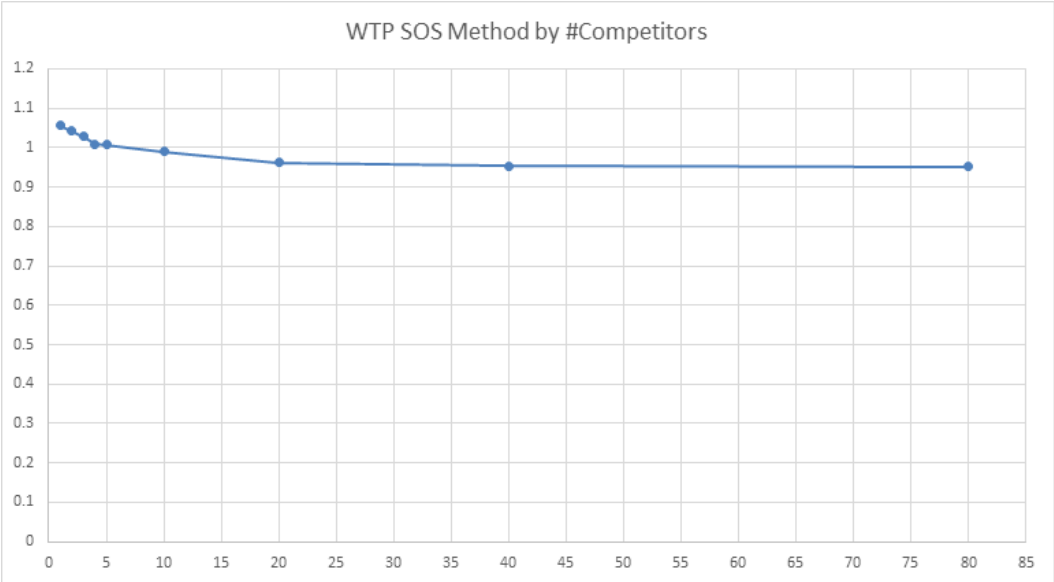
Bryan Orme

APPENDIX A

Robustness of WTP to Number of Competitors

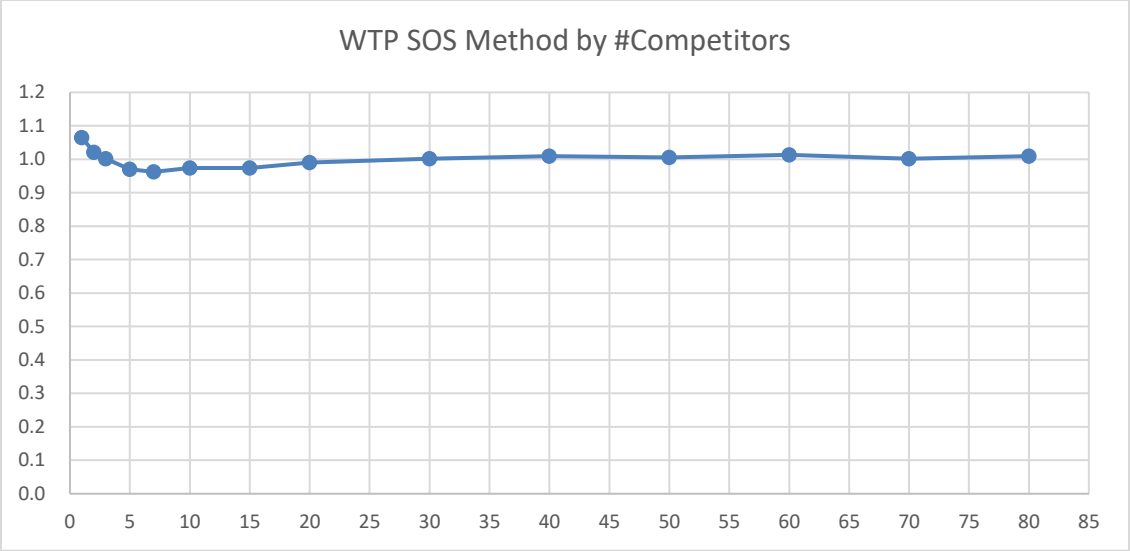
We have found that WTP estimates are fairly stable under different assumptions of number of competitors in Sampling Of Scenarios (SOS).

Using one of the CBC datasets cited earlier in this paper, we examined the robustness of WTP estimates for a given attribute level (relative to a reference level). Specifically, we varied the number of randomly drawn competitors in the simulation scenarios from 1 to 80. The results are shown below, where WTP is plotted on the Y Axis and indexed to 1.0 and number of competitors for SOS is represented on the X axis:



For this CBC dataset, the WTP with just one assumed competitor is about 7% higher than the WTP when 4 or 5 competitors are assumed. After about 20 assumed competitors, the WTP stabilizes with WTP about 5% lower than the WTP we found when using 5 assumed competitors.

We examined the same issue for a second CBC dataset and found that after five assumed competitors, the WTP results stabilized and did not change much at all:



We conclude that the SOS approach to estimating WTP is fairly robust to the number of assumed competitors. We expect results will vary somewhat depending on the data set and the number of levels in the attribute for which WTP is estimated.

For the software, we have set the default number of competitors for SOS to five. This would seem to strike a good balance between robustness of WTP results and speed.

APPENDIX B

Bootstrap Sampling within the Same HB Run vs. Separate HB Runs for Each Bootstrap Sample

To develop WTP confidence intervals, we've taken the simplification of bootstrap sampling the posterior utility estimates for respondents within the same HB estimation run (estimating HB utilities just once). Yet, since HB doesn't produce purely individual-level estimation (each individual's estimates are smoothed to some degree toward other members of the population), we may be understating the sampling distribution. Formally, it would be more appropriate to estimate WTP using independent HB estimation performed on each bootstrap sample. Unfortunately, this would dramatically increase the time requirement for obtaining confidence intervals via bootstrapping, making it prohibitive for practitioners.

How much our simplified bootstrapping approach leads to too-narrow estimates of confidence intervals depends on the degree of Bayesian shrinkage (to the upper-level model) in the posterior utility estimates. The more choice tasks per respondent, the less each respondent's utilities should be affected (via Bayesian shrinkage) by the surrounding population. Thus, for CBC datasets with relatively large numbers of choice tasks per respondent, we'd expect that the sampling distribution and resulting confidence intervals will tend to be larger and more accurate than for sparse data sets with relatively few tasks per respondent.

We compared WTP (using the simulation approach versus a set of fixed competitors) confidence interval estimates for two CBC datasets for two approaches: 1) simplified approach as used in Sawtooth Software's implementation where we use bootstrap sampling among the posterior individual-level utilities within the same HB run; 2) full approach where we conduct separate HB estimations within each bootstrap sample and use those utility estimates. For the comparisons, we used plain-vanilla HB estimation with no covariates (i.e., a single population assumption in the upper model). One CBC dataset was relatively sparse, with 8 choice sets and the other had a relatively large number of choice tasks (20).

We found with the 20-task CBC data set that the simplified approach produces confidence intervals about 25% narrower than the full approach that involves re-estimating HB utilities within each bootstrap sample. For the 8-task CBC data set, the confidence intervals were about half the width of the full approach.

We conclude that the simplified approach has the tendency to understate the width of the confidence interval for sparse CBC datasets using the plain-vanilla single population assumption in the upper model. If using our simplified bootstrapping approach and if more accurate confidence intervals for WTP estimates are needed, we recommend:

- Using CBC datasets where respondents answer at least 15 choice tasks,
- Using a few high-quality covariates (related to preference) in HB estimation to capture a more disperse representation of heterogeneity in the parameter estimates.

Taking these steps will allow use of the rapid simplified bootstrapping approach implemented in Sawtooth Software's market simulator while still achieving reasonably accurate estimates of WTP confidence intervals.

We should also note that using our software's Sampling Of Scenarios (SOS) approach tends to increase the standard error and its resulting confidence bound widths compared to using a set of fixed competitors. This isn't surprising, since sampling competitors adds another source of variability in the WTP estimates. In fact, with the SOS approach, the width of the confidence bounds may be overstated in many cases. Increasing the software's setting for number of Sampling Of Scenarios within each bootstrap loop (the default is 30) will reduce the degree of overstatement of confidence bounds. As software developers, we have to strike a balance between quality of results and time to compute. The default of 30 Sampling Of Scenarios iterations within each bootstrapping loop is one such judgement call.

REFERENCES

Allenby, Greg, Jeff Brazell, John Howell, and Peter Rossi (2014), "Economic Valuation of Product Features." *Quantitative Marketing and Economics*, 12:421–456.

Chrzan, Keith and Bryan Orme (2017), "Becoming an Expert in Conjoint Analysis," Sawtooth Software.

Orme, Bryan 2001, "Assessing the Monetary Value of Attribute Levels with Conjoint Analysis: Warnings and Suggestions," Technical Paper available at: https://www.sawtoothsoftware.com/download/techpap/monetary.pdf

Orme, Bryan 2004, "Getting Started with Conjoint Analysis," Research Publishers, Inc. First Edition.

COMMENTS ON "ESTIMATING WILLINGNESS TO PAY GIVEN COMPETITION IN CONJOINT ANALYSIS"

DAVID W. LYON
AURORA MARKET MODELING, LLC

INNOVATIONS AND STRENGTHS OF ORME'S WTP APPROACH

The preceding paper by Bryan Orme presents a defensible and well-reasoned approach to answering the "willingness to pay" question from conjoint data. This is a problem familiar to many practitioners and the source of many problems (often in the form of laughably high estimates of willingness to pay) over the years. Bryan's solution is particularly satisfying because it is conceptually simple and doesn't involve much math. There is some computation involved, but in ways that many practitioners could program themselves if they had to.

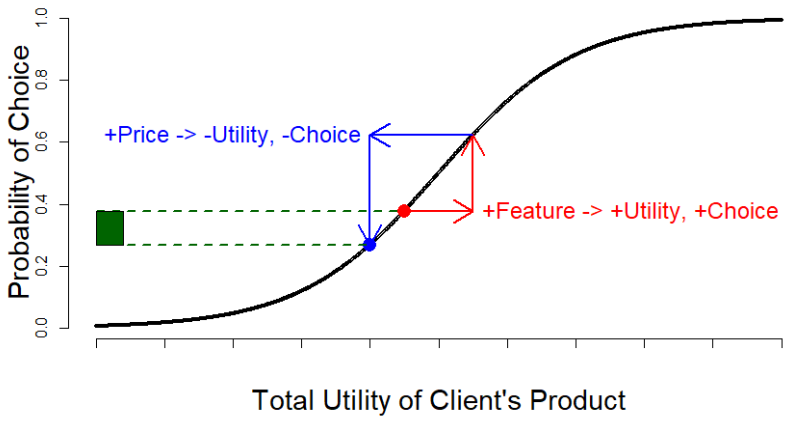
Bryan's approach recognizes or incorporates a number of important realities of WTP that some other approaches ignore:

- Respondents "on the cusp of choice" are who matter.
- WTP is not additive.
- WTP is not absolute, but dependent on the competitive context and the starting point for measurement.
- The managerial problem is to set a single price that works in the market, not to summarize the WTPs of heterogeneous respondents[6].

The Cusp of Choice

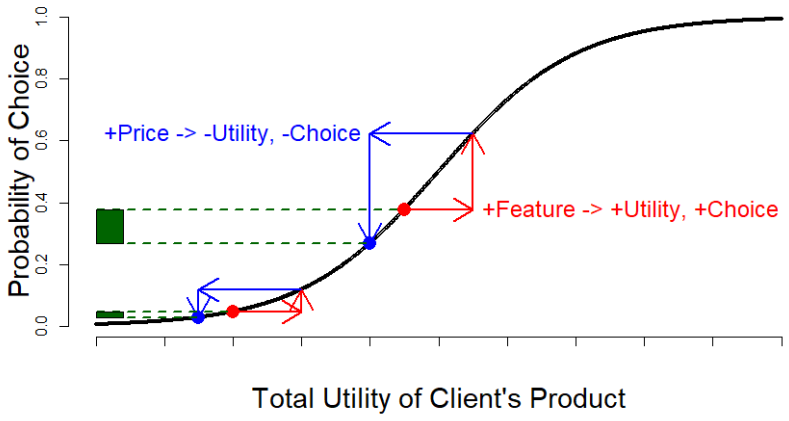
Consider the following graph of the logit curve that translates a respondent's utility for a product (x-axis) into her probability of choosing the product (y-axis), assuming some particular set of competitors.

[6] Allenby et al. (2014) address an even more on-point managerial problem: to maximize the client's profit in the context of full-knowledge competitive responses. It essentially involves setting a price that satisfies a Nash equilibrium in the "game" among competitors. This directly addresses the value of a feature, separately from WTP itself, and typically suggests a smaller price increase than any form of WTP analysis does. However, their approach places limitations on the formulation of the price utility and requires knowledge of all competitors' costs or margins.



If the respondent's utility is at the red circle on the curve and we then add a desirable feature to the client's product, utility increases and so does probability of choice, as shown by the red arrows. If we also increase the price, utility decreases, and so does probability of choice, in this case by more than the increases from the new feature. The green bar at the left highlights how much the probability of choice changed (about 10% or so). This respondent is "on the cusp of choice," or in the middle of the curve, where probability of choice is nearer 50% than to either extreme. As the steepest part of the curve, the middle is where smallish utility changes cause largish changes in choice probability.

Consider another respondent whose utilities for feature and price are identical to the first, but whose overall utility for the client product is far lower (perhaps due to a very low utility for the client's brand). In the graph below, this respondent is in the lower left corner, and even with identical price and feature utilities, his change in choice probability is far less. The respondent is very unlikely to choose the client's product in any event—they are not anywhere near "the cusp of choice."



By focusing on total simulated share, which is just the average of choice probabilities across all respondents, Bryan's approach implicitly places far more weight

on the first respondent than on the second. It does this smoothly and elegantly, without arbitrary weighting or assignment of who matters and who doesn't, but simply as a natural by-product of the logit model analysis. This is a major strength of the approach.

Non-Additivity of WTP

Bryan's reminder that WTP is not additive is timely and useful. There are many stories from the 1980s of analysts telling clients to add every possible feature and triple their price and expect market success. Unfortunately, that continued well past the 1980s. This is a common and concrete special case of the context issue noted next.

Context-Dependence and Sampling Of Scenarios

WTP is not an absolute value, but depends on the client configuration taken as a starting point, and on the competitive context. In terms of the logit curves illustrated above, respondents' starting points depend on how the client is configured to start with. Even more obviously, they depend on the competition. WTP for a feature exclusive to the client will clearly be higher than if the same feature is offered by one or more competitors.

Occasionally, the right client starting point and competitive context is obvious. For example, pharmaceutical clients may have a definite target profile for the product and the competition is often well-known, with attributes that are invariant because of either chemistry or regulation. But more often, there is no natural starting or reference point for WTP or any other kind of sensitivity analysis.

The Sampling Of Scenarios (SOS) approach is an excellent way to address this situation. It removes any need to make an arbitrary choice of starting point. The Sawtooth Software implementation that allows for prohibition of some levels for some competitors makes it particularly useful in practice by providing a way to specify the fixed parts of the situation, when we know them, while still randomly sampling the uncertain parts of the competitive situation.

The basic SOS idea should prove useful in all sorts of sensitivity analyses, not just in WTP.

The Right Managerial Problem

What underlies both the cusp of choice concept and the context-dependence concept is the adoption of the viewpoint of a price-setting manager. The fundamental WTP problem is to find a price for a feature that will maintain overall market share. (This could of course be generalized to such ideas as finding a price that would produce an x% gain in market share.)

Simulating the effects of adding a feature and increasing the price is the direct way to address that managerial problem. Too many other approaches are essentially statistical summaries, trying to find some useful way to make use of individual-level WTP calculations. But why do we care about things like algebraic equalization for those who won't buy anyway, or who will (almost) always buy? Why do we care *who* buys,

except to the extent it leads to better aggregate simulation results? For WTP purposes, we shouldn't!

Starting with the business problem and working from there is usually a good idea, and Bryan's approach does exactly that.

BOOTSTRAPPING FOR CONFIDENCE INTERVALS? WHY NOT THE HB POSTERIOR?

It is good to see confidence intervals (CIs) being explicitly considered. Too often, they are ignored until a client asks (and too few do) and then dismissed with worries about the difficulty of calculating them. It is commendable that Bryan and Sawtooth Software incorporated CI calculations from the beginning.

But, is bootstrapping the right way to get CIs? It is a solid technique, but as used here it accounts *only* for the respondent sampling variability in the results. The uncertainty in each respondent's utilities (some might say *within* each respondent's utilities) is ignored when bootstrapping on posterior means. The fundamental result of any Bayesian analysis is the posterior distribution. In our case, the HB posteriors tell us how well-determined each respondent's utility estimates are. This information is being ignored when we use just the posterior means. Bootstrapping on the posterior means does nothing to restore that lost information[7].

Instead, we could run the share-equalizing price search at the heart of Bryan's method for each of, let's say, 1000 random draws from the HB posteriors and use their variance to generate the confidence intervals. Doing so would involve no more computation than working with 1000 bootstrap samples, so the workload and timings would be the same.

Working from the HB draws, however, would incorporate *all* the sources of uncertainty in the model, not just the respondent-sampling variance. Using the draws would be much more in the spirit of Bayesian analysis, where "the posterior answers all questions," as opposed to using bootstrapping as an add-on after simplifying the hierarchical Bayes analysis down to posterior means.

In my own experiments with a single study (with 12 tasks per respondent, 10 model parameters, no covariates, and 711 respondents), the CIs for 6 different feature changes estimated from HB draws[8] were as much as 2.6 times as wide as those estimated from bootstrapping, depending on the feature, the exact method of CI construction, and whether SOS was used, with most being at least 50% wider. These are substantial differences, implying that the bootstrapping process tends to yield confidence intervals that are far too optimistic.

---

[7] As Bryan also notes, the implementation skips the HB re-estimation for each replicate that an ideal bootstrap implementation would involve. There are good practical reasons for that, but he sees as much as 25% to 50% of the total variance from the ideal being "lost" because of that simplification.

[8] This commenter used draws from the lower-level model. The upper-level draws could be used in very similar fashion.

In short, I believe the use of HB draws to replace the bootstrapping should be seriously considered. The results would be more rigorous, and the computational effort not much different[9].

Confidence Intervals with Sampling Of Scenarios

The SOS process adds additional variance (thus, widens the confidence interval) because of the variation in results for different scenarios. The current Sawtooth Software implementation averages results for 30 (or so) scenarios for each of 1000 (or so) bootstrap replications, and then calculates CIs based on the variance of those 1000 averages.

This unfortunately ignores most of the variance added by SOS. A better approach would be to calculate the variance of the 30 SOS results for each bootstrap replicate, average those variances over the 1000 bootstraps and then add the net result to the overall variance estimated from the bootstrap means. The variance of the means is the inherent respondent/model variance (not controllable once data is collected); the average variance within a replicate is that due to SOS, which can be decreased, if felt to be too large, by re-running with more than the default 30 scenario samples.

In this commenter's experiments (on the same study mentioned in the previous subsection), accounting for the variance due to SOS with 30 scenarios increased standard errors and confidence interval widths by 2% to 22%, with many values around 12%. This seems reasonably small and very practical. (Those who find it too large could cut the increase in half by using four times as many SOS samples, which would correspondingly take four times as long to run.) But it really should be accounted for.

If CIs are calculated from HB draws as suggested in the foregoing subsection, exactly parallel issues apply for SOS: the SOS variance should be accounted for, but that can be done in computationally equivalent ways.

The variance added by SOS is a smaller issue than the question of bootstrapping vs. HB draws, but it would be easy (and appropriate) to account for in any future software update.

David W. Lyon

[9] With current Sawtooth Software programs, there may be somewhat more human effort in using HB draws, in that the draws must be captured and managed, which is not common in everyday practice.

REFERENCE

Allenby, Greg, Jeff Brazell, John Howell, and Peter Rossi (2014), "Economic Valuation of Product Features." *Quantitative Marketing and Economics*, 12:421–456.