# How Should We Measure Consumers' Willingness to Pay?

# An Empirical Comparison of State-of-the-Art Approaches

KLAUS M. MILLER, RETO HOFSTETTER, HARLEY KROHMER, and Z. JOHN ZHANG

## Web Appendix A:

## Literature Overview

Table A1

STUDIES COMPARING DIRECT HYPOTHETICAL AND INCENTIVE-ALIGNED APPROACHES TO MEASURE CONSUMERS' WTP

| *Study* | *Compared Methods* | *Stimulus* | *Sample* | *Comparison Standard* | | | |
|---|---|---|---|---|---|---|---|
| | | | | *Out-of-sample choice*[a] | *Mean WTP* | *Demand Curve* | *Pricing Decision* |
| Backhaus,Wilken, Voeth, and Sichtmann (2005) | OE vs. IOE[b] | Weekend trip | Graduate students (n = 434) | n.a. | OE > IOE* | OE > IOE* | n.a. |
| Ding, Grewal, Liechty (2005) | OE vs. BDM | Chinese dinner special | Students (n = 108) | OE 7% BDM 15%[n.p.] | n.a. | n.a. | n.a. |
| Kaas and Ruprecht (2006) | OE vs. BDM | Chocolate bar | Students (n = 50) | n.a. | *OE > BDM*** | n.a. | n.a. |
| Silva, Nayga, Campbell, and Park (2007) | OE vs. BDM | Grapefruits | Grocery store shop-pers (n = 245) | n.a. | *OE > BDM** | *OE > BDM** | n.a. |
| Wertenbroch and Skiera (2002) | OE vs. BDM | Coke can, pound cake | Beach visitors (n = 200), Ferry passengers (n = 200) | n.a. | OE > BDM*** | OE > BDM[n.p.] | n.a. |

** p < .10, ** p < .05, *** p < .01, n.p. = p-value not available, n.a. = not available*
[a] Percentage of correctly predicted choices in a holdout sample.
[b] The study compared direct open-ended question-formats with (incentive OE) and without buying obligation (OE). The buying obligation was implemented by recording respondents' names and telephone numbers. However, we are unsure if the buying obligation was actually carried out.

# Table A2

## STUDIES COMPARING INDIRECT HYPOTHETICAL AND INCENTIVE-ALIGNED APPROACHES TO MEASURE CONSUMERS' WTP

| *Study* | *Compared Methods* | *Stimulus* | *Sample* | *Comparison Standard* | | | |
|---|---|---|---|---|---|---|---|
| | | | | *Out-of-sample choice*[a] | *Mean WTP* | *Demand Curve* | *Pricing Decision* |
| Backhaus,Wilken, Voeth, and Sichtmann (2005) | LCA vs. ILCA[b] | Weekend trip | Graduate students (n = 434) | n.a. | LCA ≠ ILCA[n.s.] | LCA≠ILCA[n.s.] | n.a. |
| Ding, Grewal, Liechty (2005) | CBC vs. ICBC | Chinese dinner special, snack combo | Students (n = 108, n = 52) | CBC 26% ICBC 48%[n.p.] CA 13%, IACA 18%[n.p.] | n.a. | n.a. | n.a. |
| Ding (2007) | CBC vs. ICBC | Apple iPod & accessories | Students (n =49, n = 117) | CA 17%, IACA 36%* CA 21%, IACA 34%* | n.a. | n.a. | n.a. |
| Lusk and Schroeder (2004) | CBC vs. ICBC | Beef steak | Random consumers (n = 104) | n.a. | CBC > ICBC* | CBC>ICBC[n.p.] | n.a. |
| Silva, Nayga, Campbell, and Park (2007) | CBC vs. ICBC | Grapefruits | Grocery store shop-pers (n = 245) | n.a. | CBC > ICBC* | CBC > ICBC* | n.a. |
| Voelckner (2006) | LCA vs. ILCA[c] | Prepaid phone cards | Students and university employees (n = 1089) | n.a. | LCA > ILCA* | CA > ILCA* | n.a. |

** p < .10, ** p < .05, *** p < .01, n.s. = not significant, n.p. = no p-value available, n.a. = not available*

[a] Percentage of correctly predicted choices in a holdout sample.

[b] The study compared so-called limit conjoint analysis with (ILCA)and without (LCA) buying obligation. The buying obligation was implemented by recording respondent's name and telephone number. However, we are unsure if the buying obligation was actually carried out.

[c] The study compared so-called limit conjoint analysis with (ILCA) and without (LCA) buying obligation. It becomes not clear from the study whether the participants actually bought the product after completion of the research.

# Table A3

## STUDIES COMPARING DIRECT AND INDIRECT APPROACHES TO MEASURE CONSUMERS' WTP

| *Study* | *Compared Methods* | *Stimulus* | *Sample* | *Comparison Standard* | | | |
|---|---|---|---|---|---|---|---|
| | | | | *Out-of-sample choice*[a] | *Mean WTP* | *Demand Curve* | *Pricing Decision* |
| Backhaus, Wilken, Voeth, and Sichtmann (2005) | OE vs. LCA[b]<br>IOE vs. ILCA[c] | Weekend trip | Students (n = 434) | n.a. | OE < LCA$^{***}$<br>IOE < ILCA$^{***}$ | OE < LCA$^{n.p.}$<br>IOE < ILCA$^{***}$ | n.a. |
| Ding, Grewal, Liechty (2005) | OE vs. CBC<br>BDM vs. ICBC | Chinese dinner special | Students (n = 08) | OE 7%<br>CBC 26%$^{n.p.}$<br><br>BDM 15%<br>ICBC 41%$^{n.p.}$ | n.a. | n.a. | n.a. |
| Kalish and Nelson (1991) | OE vs. CA[d] | Airlines ticket | Students (n = 255) | OE 46%<br>CA 62%$^{n.p.}$ | n.a. | n.a. | n.a. |
| Lusk and Schroeder (2006) | BDM vs. ICBC | Beef steaks | Random consumers (n = 151) | n.a. | BDM < ICBC$^{n.p.}$ | n.a. | n.a. |
| Silva, Nayga, Campbell, Park (2007) | OE vs. CBC<br>BDM vs. ICBC | Grapefruits | Grocery store shoppers (n = 245) | n.a. | OE > CBC$^{*}$<br>BDM > ICBC$^{*}$ | OE > CBC$^{*}$<br>BDM > ICBC$^{*}$ | n.a. |
| Veisten (2007) | OE vs. CBC | Furniture | Furniture shoppers (n = 590) | n.a. | OE < CBC$^{***}$<br>OE > CBC$^{***}$ | n.a. | n.a. |
| Voelckner (2006) | OE vs. LCA[b] | Prepaid phone card | Students/ university employees (n = 1089) | n.a. | OE > LCA$^{***}$<br>OE < LCA$^{***}$ | OE > LCA$^{***}$<br>OE < LCA$^{***}$ | n.a. |

** p < .10, ** p < .05, *** p < .01, n.s. = not significant, n.p. = no p-value available, n.a. = not available*

[a] Percentage of correctly predicted choices in a holdout sample.

[b] The study compared an open-ended question format (OE) to so-called limit conjoint analysis (LCA)

[c] The study compared an open-ended question format with buying obligation (incentive OE or IOE) to so-called limit conjoint analysis with buying obligation (incentive LCA or ILCA). The buying obligation was implemented by recording respondent's name and telephone number. However, we are unsure if the buying obligation was actually carried out.

[d] The study used ranking and rating conjoint analyis (CA) as well as simulated choices instead of actual choices for the holdout task.

## Web Appendix B:

## Adequate Number of Choice Tasks in Our Choice-based Conjoint Analysis

Prior choice-based conjoint studies have used anywhere from three up to 32 choice tasks for parameter estimation (see for an overview Haaijer and Wedel 2001). Several prior studies used three to five choice tasks per respondent (e.g., Dellaert, Borgers, and Timmermans 1995; Hennig-Thurau et al. 2007; Oppewal, Louviere, and Timmermans 1994; Vriens, Oppewal, and Wedel 1998). In line with these studies, in our application of standard (i.e., hyp othetical) and incentive-aligned choice-based conjoint analysis (CBC and ICBC groups), we use a total of seven choice tasks per respondent. Five tasks are random and used for individual parameter estimation and two tasks are holdout tasks which we use for validation purposes.

Given the wide range in the number of choice tasks used in prior studies, it is worth asking whether our five choice tasks are enough to yield adequate utility estimates. A study by Lenk et al. (1996) found that reduced experimental designs can generate adequate results when estimating individual partworths using Hierarchical Bayes estimation for choice-based conjoint analysis. Further, Toubia et al. (2003) found little difference between eight and 16 choice tasks, but in a different context Toubia, Hauser, and Simester (2004) found large differences between five and 16 choice tasks. From the literature it is unclear how many choice tasks need to be used in choice-based conjoint studies to obtain adequate measurement validity.

To shed light on this issue, we perform two analyses. First, we use the elicited five choice-tasks from the survey to gauge the effect of the number of choice-tasks on the validity of the estimated individual partworths and the resulting WTP estimates. Second, we run a simulation to analyze the effect of using more than five choice-tasks in our study.

***Analysis of the five elicited choice task in our studies:***

In order to gauge the optimal number of tasks in our specific datasets, we analyze how well a holdout task (the sixth choice task) can be predicted based on estimates using 1, 2, 3, 4, or 5 choice tasks (see Hoogerbrugge and van der Wagt 2006). We calculate hit rates and mean average errors (MAEs; see Moore, Gray-Lee, and Louviere 1998) for all cases and check for a significant improvement in estimation by adding additional choice tasks[1].

Our results indicate that in our application five choice tasks yield adequate results which can not be significantly improved by adding an additional choice task. First, hit rates show no significant improvement by moving from four to five tasks in all cases (Figure A1, Table A1). For choice-based conjoint analysis in study 1 (CBC-1), hit rates improve significantly from one to two choice tasks (z = -3.093, *p = .002*) but not from two to three tasks (z = -1.125, *p = .261*), from three to four tasks (z = -.645, *p = .519*) or from four to five tasks (z = -.162, *p = .872*). For incentive aligned choice-based conjoint in study 1 (ICBC-1), hit rates improve significantly from one to two tasks (z = -1.982, *p = .048*) but not from two to three tasks (z = -1.098, *p = .272*), from three to four tasks (z = -.252, *p = .801*) or from four to five tasks (z = .623, *p = .533*). In study 2[2], hit rates for choice-based conjoint analysis (CBC-2) are the same in the case of one and two choice tasks. Moving from two to three choice tasks significantly improves the hit rate (z = -2.170, *p = .030*). Moving from three to four (z = -.100, *p = .920*) and from four to five (z = -.606, *p = .545*) does not significantly improve the hit rate.

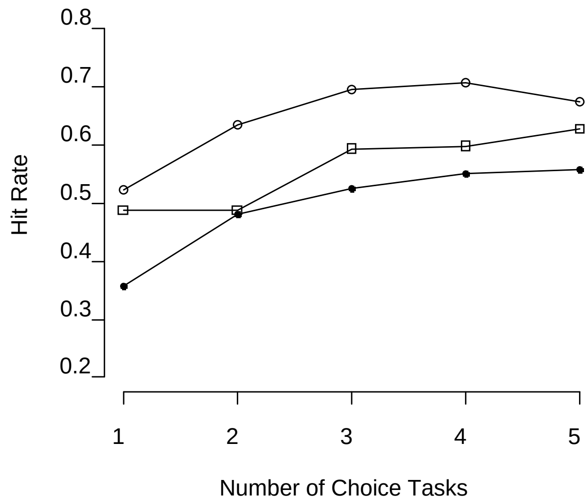
Second, mean average errors (MAEs) might further decrease in the case of CBC-1 and ICBC-1 by moving from 5 to more choice tasks (see Figure A2). This is not the case for CBC-2 where MAE is already at a minimum when 3 choice tasks are used for the estimation.

---

[1] We include two measures of out-of-sample fit as hit rates are sensitive to the reliability of individual estimates, whereas MAEs, by aggregating across individual estimates, are less affected by reliability and more by bias (Wittink and Bergestuen 1999). Thus, the two measures relate to different aspects of predictive validity.
[2] See Web Appendix F.

Figure B1

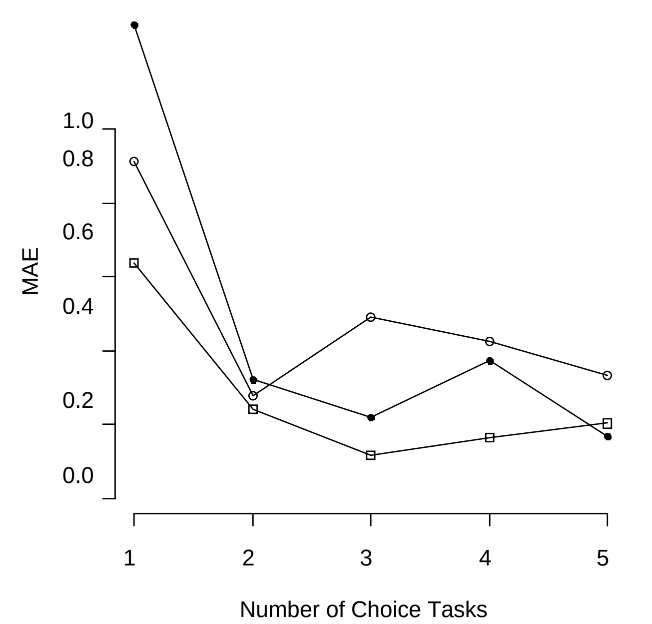
CBC CHOICE TASKS AND CORRESPONDING HIT RATES




Black dots: Choice-Based Conjoint (CBC-1)
White dots: Incentive-aligned Choice-Based Conjoint (ICBC-1)
White squares: Choice-Based Conjoint (CBC-2)

Figure B2

CBC CHOICE TASKS AND CORRESPONDING MAES




Black dots: Choice-Based Conjoint (CBC-1)
White dots: Incentive-aligned Choice-Based Conjoint (ICBC-1)
White squares: Choice-Based Conjoint of the second study (CBC-2)

Table B1

HIT RATES AND MAES PER CBC SETTING

| *Number of Choice Tasks* | *CBC-1* | | *ICBC-1* | | *CBC-2* | |
|---|---|---|---|---|---|---|
| | *Hit Rate* | *MAE* | *Hit Rate* | *MAE* | *Hit Rate* | *MAE* |
| 1 | .358 | 1.284 | .523 | .914 | .488 | .638 |
| 2 | .481 | .323 | .636 | .278 | .488 | .242 |
| 3 | .526 | .219 | .695 | .490 | .594 | .116 |
| 4 | .552 | .374 | .709 | .424 | .599 | .164 |
| 5 | .558 | .168 | .676 | .331 | .628 | .203 |

Note: The chance model for all CBC and ICBC applications has a hit rate of 20% (four products plus none option in holdout-task)
Note: Hit rates seem comparable to other studies such as Ding, Grewal, and Liechty (2005) who reported a 48% hit rate.

As it is our aim to measure consumers' maximum WTP, we are also interested how the number of choice tasks influences these values. For this purpose, we run a sensitivity analysis to analyze the impact of the number of choice tasks on the resulting WTP estimates. Specifically, we estimate individual partworths three times, based on three, five, and seven choice tasks, and then use the resulting partworths to calculate individual level WTP values (see the method section in the paper on how this calculation is carried out). This leaves us with three sets of WTP estimates for the same group of respondents (see Figure A3 for a plot of the resulting demand curves).

In our dataset, we elicit a total of seven choice tasks from each individual respondent. This allows us to consider a subset of 7 - x choice tasks for utility estimation. In the case of three choice tasks, we select the first three choice tasks which the respondent has answered. While answering the first three tasks, the respondent did not know how many more choice tasks would follow. Hence, we can make a cut after the first three choice tasks and only used these for further analysis. The same argument applies in the case of five choice tasks. In the case of seven choice tasks, we use all available choice tasks. We apply this analysis to the CBC-1, ICBC-1 and the CBC-2 dataset.

In the CBC-1 dataset, we find that WTP values stemming from the three and five choice tasks datasets differ significantly in terms of their mean (Welch two sample t-test: $t = -3.727$, $p = .000$) and their distribution (Two-sample Kolmogorov-Smirnov test: $D = .281$, $p = .000$). This indicates a significant difference in the resulting WTP values when choosing five over three choice tasks. However, if we compare WTP values stemming from five and seven choice tasks datasets, no significant difference in the WTP means ($t = 1.376$, $p = .169$) and the distributions ($D = .068$, $p = .444$) can be found. This indicates that in our application there is no significant difference in the resulting WTP values no matter if we choose five or seven choice tasks.

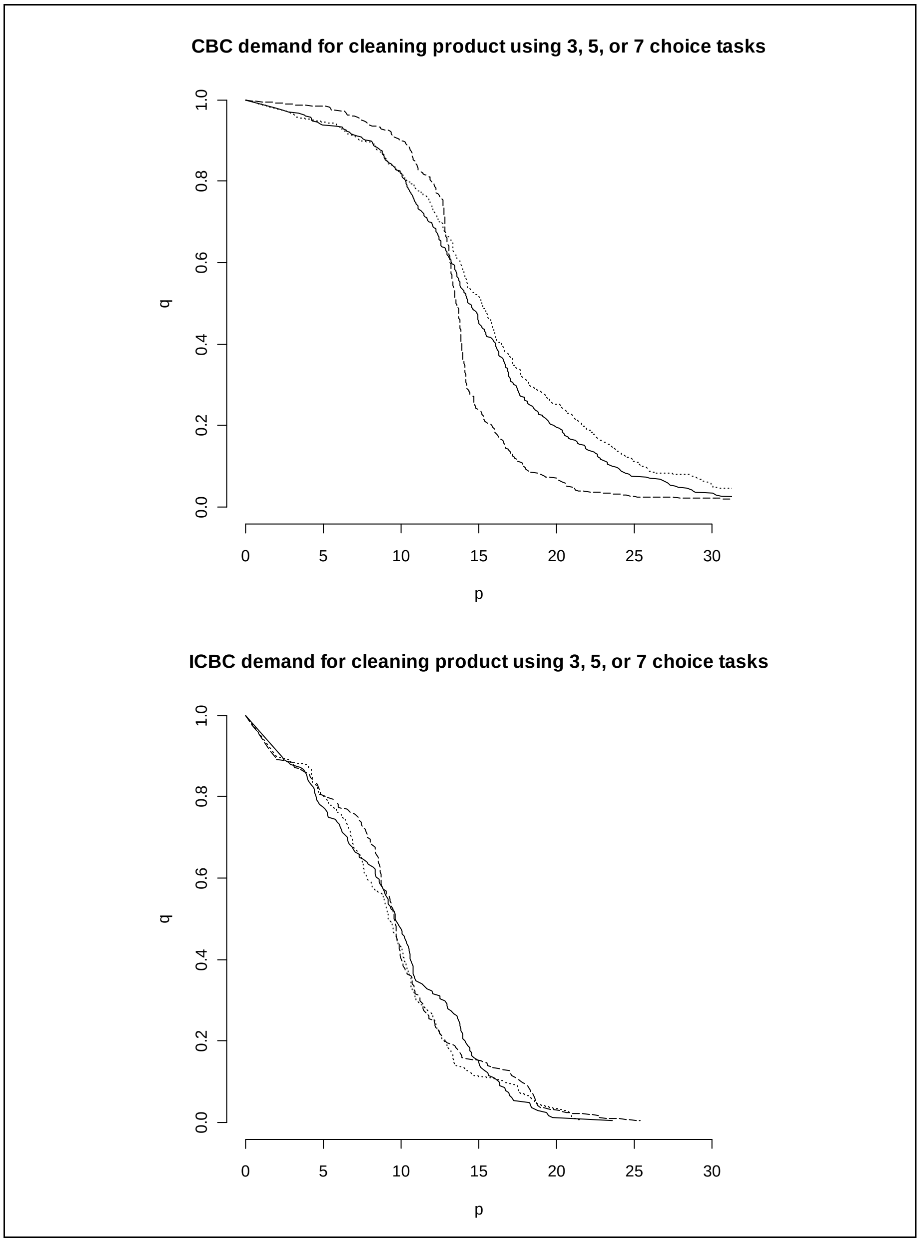
In the ICBC-1 dataset, we find no significant differences when moving from three to five choice tasks (t = .689, *p = .491*; D = .110, *p = .277*) as well as when moving from five to seven (t = 1.376, *p = .169*; D = .068, *p = .444*) choice tasks datasets (see Figure A4).

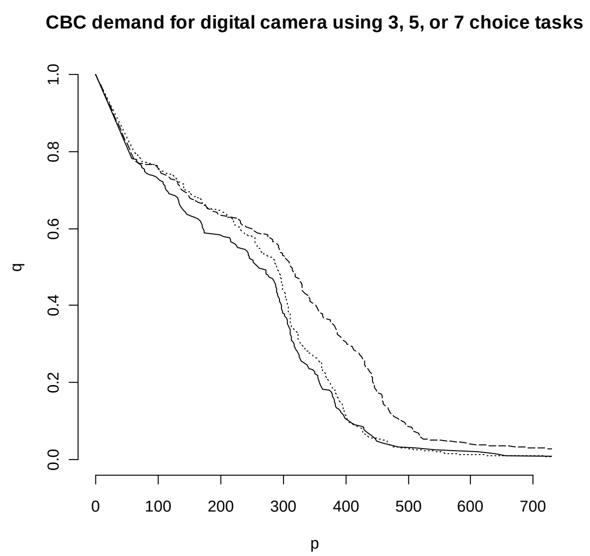
We obtain a similar result in the CBC-2 dataset: Mean WTP and distribution of WTP values differ significantly by moving from three to five choice tasks (t = 2.081, *p = .038*, D = .203, *p = .004*). However, we find no significant difference on both tests when moving from five to seven choice tasks (t = .950, *p = .343*, D = .077, *p = .567*)

Our analysis indicates that WTP values will be significantly different if we choose five over three choice tasks for individual-level parameter estimation. The results, however, will not be significantly different if we choose seven over five choice tasks. Hence, choosing sev-en over five choice tasks would not have affected the resulting WTP values significantly in all cases of our study.

Figure B3

ESTIMATED CBC DEMAND USING 3, 5, OR 7 CHOICE TASKS






Dashed line: estimated demand based on 3-choice tasks
Dotted line: estimated demand based on 5-choice tasks
Solid line: estimated demand based on 7-choice tasks

***Simulation:***

In the previous analysis, of course, it remains unclear what happens if the number of choice tasks is expanded to more than seven choice tasks. For this analysis, we perform a simulation as follows: We use estimated partworths from study 1 and treat them as real data, assuming these partworths will correctly reflect the partworths of real consumers. Using these partworths as input for choice decisions, we simulate choices in 19 CBC settings that differ only in the number of choice tasks. The profiles shown (including the none-choice option), the attributes, and the attribute levels were all equal to the ones in study 1. We use a computer-generated design with complete enumeration and full randomization. We apply maximum utility and first choice as choice rules in each task. We generate choice data for 300 respondents in each of the 19 CBC settings. In each setting, we simulate choices for x (1 to 19) random tasks and one holdout task. Only the random tasks were used for the estimation of individual partworths in each setting. We estimate individual partworths using the Hierarchical

Bayes routine in Sawtooth's CBC/HB module with 10,000 iterations (after 10'000 burn-in iterations).

We evaluate the estimated partworths using three measures (see Table A2 and Figures A4, A5, and A6), mean average error (MAE), correlation of real partworths with estimated partworths, and hit rates. We find that MAEs are strongly reduced by moving from one to five choice tasks, but only minor reductions in MAE are possible by moving from five to more choice tasks. Correlations also improve from .879 to .952 (+8.03%***[3]) by moving from one to five choice tasks. Correlations still improve significantly but only little when moving from five to 19 choice tasks (.952 to .968, +1.64%***). Hit-Rates improve significantly by moving from one to five (z = -4.840, $p < .001$[4]) and from five to 19 choice tasks (z = -2.353, $p = .019$). Hit-rates do not differ significantly when moving from five to ten (z = -.776, $p = .438$) or 15 choice tasks (z = -1.123, $p = .261$).

The results of the simulation suggest that in our case, increasing the number of choice tasks to more than five would not have resulted in much gain in terms of hit rates, MAEs and the correlation between real and estimated partworths.

The improvement in correlations between real and estimated partworths is significant, but only a little (+1.64%), when moving from five to 19 choice tasks. Hit-rates are not improved significantly by moving from five to ten or 15 choice tasks. Only moving from five to 19 choice tasks would significantly improve the hit-rate. However, it can be questioned if moving from five to 19 choice tasks would bring an improvement in real survey settings as the respondents' answers might suffer from a response bias in longer conjoint settings (e.g., due to respondent fatigue, see e.g., Lenk et al. 1996). In our simulation, we do not account for such response biases.

---

[3] Blalock (1972).
[4] Keller, Warrack, and Bartel (1990).

To sum up, as the results of the simulation and the analysis of the datasets from study 1 and study 2 suggest, moving from five to more choice tasks might not have yielded large improvements in the validity of individual partworths and hence, individual WTP measures. Given the fact that the bias of hypothetical CBC is substantial in both studies, it is unlikely that marginal improvements in the validity of the partworths estimates would have changed the results of the method comparison much. Further, in the incentive aligned condition, ICBC is capable to capture the real WTP in terms of mean, distribution, and the outcome of a pricing decision, which is another indicator of the validity of our choice-based conjoint design using five choice tasks per respondent. We conclude that in our case, using five choice tasks for individual partworth estimation yields acceptable results.

However, we acknowledge the fact that the number of choice tasks used under CBC/ICBC may influence parameter estimation and the conclusions of the method comparison. The simulation gave us some insights concerning the sensitivity of the results if we had used a larger number of choice tasks per respondent, but we do not have the empirical data for this case. Hence, we can not rule out the alternative explanation unambiguously and do not know whether our results would hold if we had used a larger number of choice tasks per respondent. Future researchers could replicate our findings with a larger number of choice tasks and investigate this issue.

Table B2

NUMBER OF CHOICE TASKS USED FOR PARTWORTH ESTIMATION AND ACCORDING HIT-RATES, MAES, AND CORRELATIONS BETWEEN REAL AND ESTIMATED PARTWORTHS (SIMULATION)

| *Number of choice tasks used for estimation* | *Hit-Rates* | *MAEs* | *Correlation between real and estimated partworths* |
|---|---|---|---|
| 1 | .653 | .640 | .879 |
| 2 | .690 | .467 | .916 |
| 3 | .693 | .513 | .928 |
| 4 | .763 | .313 | .946 |
| 5 | .827 | .127 | .952 |
| 6 | .823 | .153 | .954 |
| 7 | .820 | .220 | .948 |
| 8 | .833 | .093 | .956 |
| 9 | .853 | .067 | .953 |
| 10 | .850 | .127 | .959 |
| 11 | .867 | .100 | .958 |
| 12 | .863 | .053 | .968 |
| 13 | .857 | .060 | .965 |
| 14 | .850 | .087 | .970 |
| 15 | .860 | .033 | .938 |
| 16 | .867 | .053 | .966 |
| 17 | .893 | .060 | .965 |
| 18 | .893 | .060 | .965 |
| 19 | .893 | .060 | .968 |

Figure B4

CBC CHOICE TASKS AND CORRESPONDING MAES (SIMULATION)

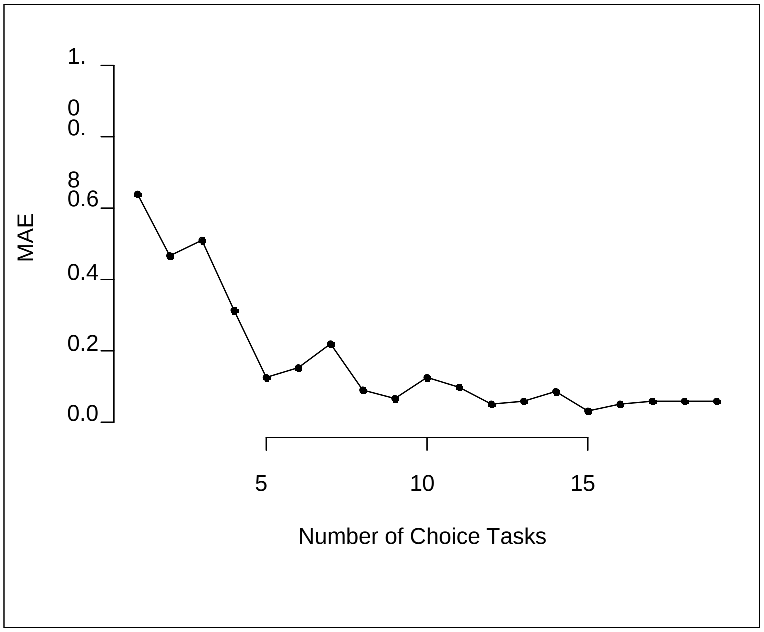



Figure B5

CORRELATIONS BETWEEN REAL AND ESTIMATED UTILITIES (SIMULATION)

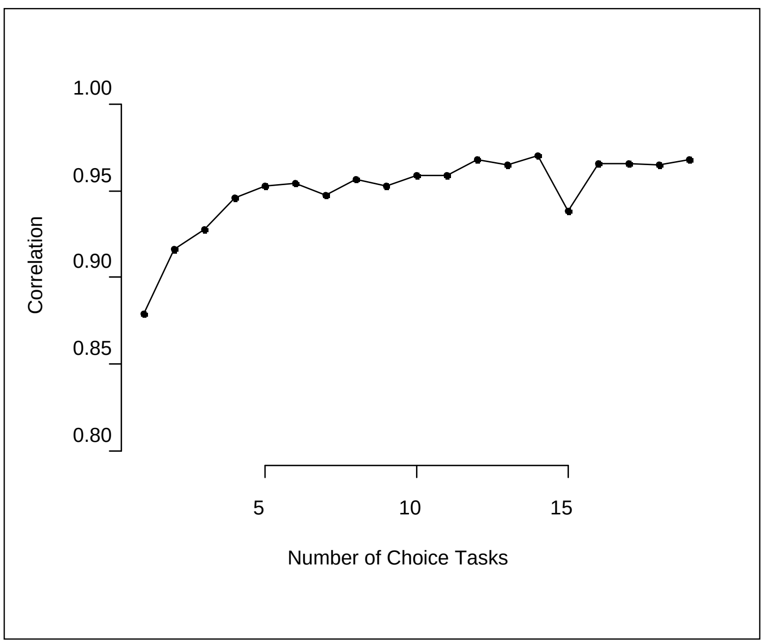
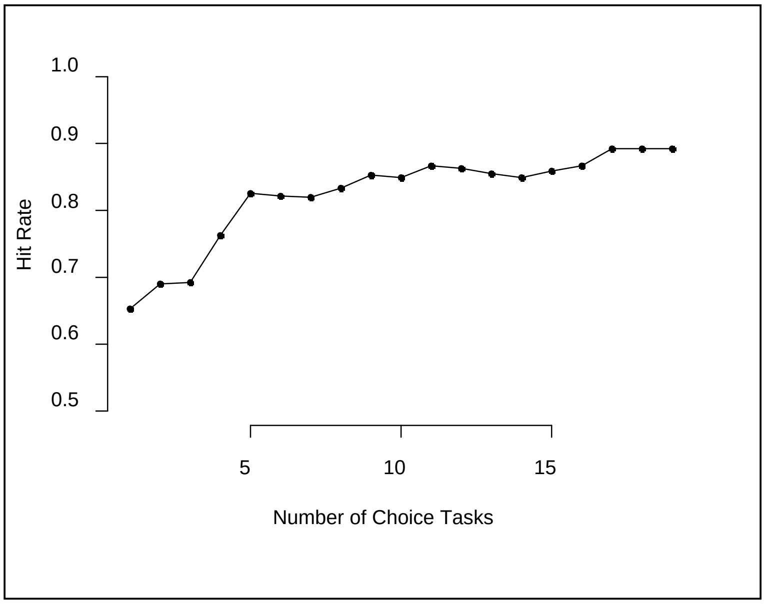


Figure B6

CBC CHOICE TASKS AND CORRESPONDING HIT RATES (SIMULATION)



# Web Appendix C:

## Exploratory Comparison of Linear and Piece-Wise Linear Approach to Calculate Consumers' WTP

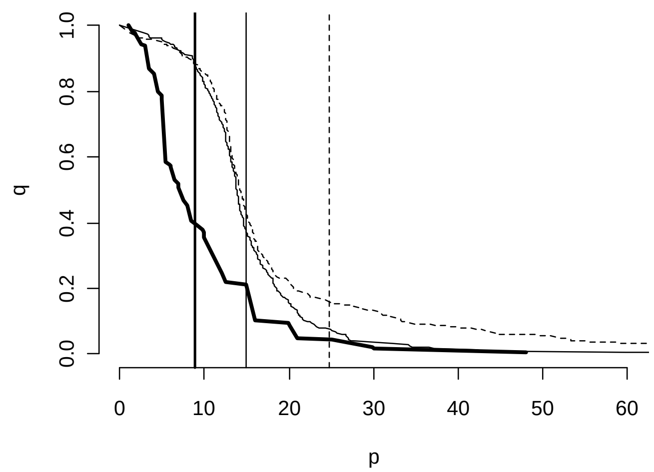
Traditionally, a single linear price parameter is estimated in choice-based conjoint analysis (referred to as "linear approach" in the remainder) which is then used in calculating consumers' WTP (see e.g., Jedidi and Jagpal 2009). This approach has been found to yield extremely fat tails in posterior WTP distributions (Sonnier, Ainslie, and Otter 2007). In the present paper, we apply a piece-wise linear approach and use multiple linear price parameter estimates for WTP calculation (referred to as "piece-wise linear" approach in the remainder).

We find that WTP values from the linear approach differ significantly from the values of the piece-wise linear approach. The piece-wise linear mean (24.746), variance (3,813.129), and distribution are significantly different from the linear mean (14.918, t = 2.7822, $p < .01$), variance (55.243, F = 69.025, $p < .01$), and distribution (D = .110, $p < .05$). Using BDM as a benchmark, WTP distributions of both approaches deviate significantly (linear: D = .606, $p < .01$; piece-wise linear: D = .577, $p < .01$). There is less deviation from the benchmark, however, in the case of the piece-wise linear approach. Graphical analysis confirms those findings and shows the "fat tails" in the demand curve for the linear approach (Figure B1). Based on these findings, we assume that a piece-wise linear approach may be able to yield more valid WTP estimates for our case compared to the traditional linear approach.

Figure C1

Linear vs. Piece-Wise Linear Approach to Calculate

Individual WTP from Choice-Based Conjoint Data




Fat curve: BDM,
Normal curve: CBC, piece-wise linear approach
Dashed curve: CBC, linear approach
Vertical lines: Mean values

## Web Appendix D:

## Numerical Example for Piece-Wise Linear WTP Calculation

The following numerical example demonstrates the piece-wise linear WTP calculation: consider a respondent with a none-choice utility of $u^* = 4.158$ and $u_{it|\sim p} = 1.58$. Assume that partworths for 13 price levels are available ranging from $u(p_1) = -8.53$ (highest price) to $u(p_{13}) = 6.82$ (lowest price). Looking for $\min\left(u(p_j)\right)$ that satisfies equation (1) yields $u(p_{10}) = 3.01$ such that equation (1) is satisfied:

$$1.58 + 3.01 \geq 4.158$$

Hence, $p_{10}$ represents the price level that is the closest to the respondents' maximum WTP while still satisfying equation (1). In this case, consuming the product at price $p_{10}$ results in a higher total utility than not consuming it (none-choice utility). The next higher price $u(p_9) = 2.08$ would not satisfy equation (1) any more. However, the product's total utility at $p_{10}$ is still higher than the utility of the non-choice option, indicating that the respondent would be willing to pay more than $p_{10}$ but not as much as $p_9$. Plugging the values into equation (2) yields $u_i^* - u_{it|\sim p} = 2.578$ . Calculating the inverse utility function (using linear interpolation) then yields WTP = 4.82, which is maximum WTP in this case.

## Web Appendix E:

## Sensitivity of p* to the parameters of the logit model, alpha and beta

In a monopoly, calculation of the profit-maximizing price p* is straightforward if costs and demand are known. p*, of course, is sensitive to changes in the underlying demand curve used in such calculation. In the present paper, a logit model of the following form is used to model demand (e.g., Wertenbroch and Skiera 2002):

$$quantity(price) = \frac{e^{\alpha+\beta\times price}}{1+e^{\alpha+\beta\times price}}$$

As we will show in the following section, p* is sensitive to changes in both parameters alpha and beta of such a logit model.

Monopoly profit can be calculated as follows (assuming no fix costs[5] and constant variable costs cost information according to cleaning product manufacturer):

$$profit(price) = quantity(price)\ \ (price - costs)$$

To find the profit maximizing price we set the first derivative of the profit function equal to zero

$$0 = \frac{\beta e^{\alpha+\beta\times price}(price - costs)}{1+e^{\alpha+\beta\times price}} - \frac{(e^{\alpha+\beta\times price})^2 (price - costs)\beta}{(1+e^{\alpha+\beta\times price})^2} + \frac{e^{\alpha+\beta\times price}}{1+e^{\alpha+\beta\times price}}$$

and solve for the profit maximizing price, p*.

$$p^* = \frac{\beta \times \text{costs} - 1 - e^{-LambertW(e^{\beta\times\cos ts - 1+\alpha}) + \beta\times\cos ts - 1 + \alpha}}{\beta}$$

The marginal sensitivity of p* to beta and alpha is then given as follows:

---

[5] fix costs cancel out in the derivation.

$$\frac{dp^*}{d\beta} = \frac{1}{\beta}\left( costs - \left( -\frac{costs \times LambertW(e^{\beta \times \cos ts - 1 + \alpha})}{1 + LambertW(e^{\beta \times \cos ts - 1 + \alpha})} + costs \right) e^{-LambertW(e^{\beta \times \cos ts - 1 + \alpha}) + \beta \times \cos ts - 1 + \alpha} \right)$$
$$- \frac{\beta \times costs - 1 + e^{-LambertW(e^{\beta \times \cos ts - 1 + \alpha}) + \beta \times \cos ts - 1 + \alpha}}{\beta^2}$$

$$\frac{dp^*}{d\alpha} = \frac{1}{\beta}\left( \left( -\frac{LambertW\left(e^{\beta \times \cos ts - 1 + \alpha}\right)}{1 + LambertW\left(e^{\beta \times \cos ts - 1 + \alpha}\right)} + 1 \right) e^{-LambertW\left(e^{\beta \times \cos ts - 1 + \alpha}\right) + \beta \times \cos ts - 1 + \alpha} \right)$$

It is easily observable that both marginal effects are dependent on both alpha and beta. We can thus not say that p* is sensitive to either beta or alpha alone.

Additionally, we are interested in how variations in alpha and beta affect our estimate of p* in a real pricing situation. For this purpose, we calculate the effects of marginal changes in alpha and beta on p* in the case of the real demand curve (see REAL demand in the paper) by plugging in the values of Table 4 for REAL (alpha = 1.929, beta = -.244) and the value for the constant marginal costs cost = .85 (see paper p. 19). The absolute marginal effects are as follows:

$$\frac{dp^*}{d\alpha} = 1.902$$

$$\frac{dp^*}{d\beta} = 32.955$$

If we increase alpha by one unit, p* will increase by 1.902. If we increase beta by one unit, p* will increase by 32.955. Based on the absolute marginal effects, we can calculate the relative sensitivity of p* to changes in alpha and beta. Based on REAL demand, if alpha changes by 1%, p* changes by .432%. If beta changes by 1%, p* changes by .946%.

If we plug in these relative sensitivities, we can calculate the % change in p* due to changes in alpha and beta for all demand models (see Table D1).

Table E1

SENSITIVITY OF P* TO CHANGES IN LOGIT MODEL PARAMETERS

| *Method* | *α* | *β* | *% Change in α* | *% Change in β* | *% Change in p* due to % change α* | *% Change in p* due to change in β* | *% Change in p* due to change in α and β* | *Approximation for p** |
|---|---|---|---|---|---|---|---|---|
| OE | 2.888 | -.267 | 49.715 | 9.426 | 21.456 | -8.917 | 12.539 | 9.566 |
| CBC | 4.569 | -.3208 | 136.858 | 31.475 | 59.066 | -29.776 | 29.290 | 10.990 |
| BDM | 2.420 | -.286 | 25.454 | 17.213 | 10.985 | -16.284 | -5.298 | 8.050 |
| ICBC | 3.857 | -.407 | 99.948 | 66.803 | 43.136 | -63.196 | -20.060 | 6.795 |
| REAL | 1.929 | -.244 | - | - | - | - | - | 8.500 |

We find that both hypothetical methods (OE and CBC) have a large error in alpha, but a relatively lower error in beta. This causes the curve to move outwards while approximately keeping the slope. Incentive-aligned methods (BDM and ICBC) tend to have errors in both alpha and beta. As these errors are of relatively similar magnitude, they tend to cancel out which leaves us with an overall good approximation of p*. This finding may or may not be unique to our application.

# Web Appendix F:

# Digital Camera Study (Study 2)

To test our hypothesis that CBC may perform better when a product is less unique and has more competing products, unlike our cleaning product to a preliminary test, we analyzed a second set of data. Although the data was collected the same way as for study 1, it had three important differences. First, the stimulus we used was an innovative digital compact camera (10.0 Megapixel, 4x Zoom, compact body type) supplied to us by a large multinational manufacturer of electronic equipment who planned to launch the product in Switzerland with the suggested retail price of Swiss Francs (CHF) 288.00. Second, we had three treatment conditions: OE, CBC, and BDM respectively. As both incentive-aligned methods (IACA, BDM) in study 1 were able to capture true demand, we took the BDM mechanism as a benchmark for the measurement of true demand. Third, a total of five attributes are identified with our digital camera: brand, shape/body, megapixel, optical zoom, and price (see Table F1), resulting in a $4^1 3^3 5^1$ attribute space. We gathered a total of 152 responses in the OE group, 94 in the BDM group, and 207 in the CBC group.

Our analysis shows that both OE and CBC generate significantly biased mean values (Table F2). In both methods, mean confidence intervals do not overlap with confidence intervals of the BDM (true) mean. T-tests confirm this hypothetical bias (OE-BDM: $t = 11.729$, $p < .001$; CBC-BDM: $t = 9.174$, $p < .001$). In addition, the lower hypothetical WTP/ actual WTP ratio for CBC indicates that CBC captures true mean WTP significantly better than OE. All the distributions differ significantly from each other (Table F3). Only when applying the LR test to OE and CBC, no significant difference can be found. Additionally, correlation between OE and CBC demand is high ($r = .870$, $p < .05$). These findings from this second dataset indicate that OE and CBC are significantly different from BDM, but CBC is less biased.

Finally, for the economic analysis, given that the product is not distributed directly by the manufacturer, we can use the wholesale price of CHF 207 as the marginal cost. Market size was reported to be 10,000 units for this specific camera in the target market by the manufacturer in 2008. Our analysis suggests that the optimal price is not significantly different from the benchmark in both OE and CBC (see Table F4), although the absolute difference is substantial enough. Moreover, in forecasting the optimal quantity and profits, neither OE nor CBC is able to capture the true value, although CBC is marginally more accurate than OE in this second application. Thus, we found that CBC might perform better than OE for a more expensive, less frequently purchased, durable product where competing products are available. We added Tables F5, F6, and Figure F1 for the sake of comparability to the first study.

Table F1

ATTRIBUTES AND LEVELS INCLUDED IN STUDY 2

CHOICE BASED CONJOINT ANALYSIS

| *Attribute* | *Levels* | *Number of Attribute Levels* |
|---|---|---|
| Brand | BRAND-A, BRAND-B, BRAND-C, BRAND-D | 4 |
| Megapixel | 8.0, 10.0, 14.0 | 3 |
| Zoom | 2x, 4x, 8x | 3 |
| Body | Compact type 1, Compact type 2, Semi-professional | 3 |
| Price | CHF 58, CHF 175, CHF 288, CHF 405, CHF 520 | 5 |

Table F2

MEAN WTP, STANDARD ERRORS, AND 95% CONFIDENCE INTERVALS

| *Method* | *n* | *Mean (Swiss Francs)* | *95% Confidence Interval* [d] | *Ratio HWTP or AWTP / Benchmark* |
|---|---|---|---|---|
| OE[b, a] | 152 | 292.39 | [269.888, 316.033] | 2.90 |
| CBC[c, a] | 207 | 247.38 | [223.3479, 271.1608] | 2.46 |
| BDM[b, c] (Benchmark) | 94 | 100.71 | [84.766, 132.149] | n.a. |

[a] Values with same superscripts differ at $p < .01$ in a pairwise t-test, [b] $p < .001$, [c] $p < .001$
[d] For confidence intervals of the WTP means, we apply a nonparametric approach due to the possibility of skewed WTP values. We use the bias-corrected and accelerated bootstrap percentile (BCa, 10,000 iterations, see Efron and Tibshirani 1993) method for this purpose.
Notes: n.a. = not applicable.

Table F3

TEST RESULTS COMPARING THE WTP DISTRIBUTIONS OF OE AND CBC WITH THE BDM BENCHMARK

| *Comparison* | *Pearson's r* | *Kolmogorov-Smirnov (KS) Test* | *Likelihood Ratio (LS) Test*[a] |
|---|---|---|---|
| OE – BDM | .701 | .698*** | 31.351*** |
| CBC – BDM | .604 | .589*** | 36.195*** |
| OE - CBC | .980 | .216*** | .506 |

* p < .1, ** p < .05, *** p < .001, Notes: n.a. = not applicable
[a] The Likelihood Ratio (LR) test tests the null hypothesis of equal distributions. The LR is calculated as follows (e.g. for a comparison between REAL and BDM): $LR = -2*(LL_{pooled} - (LL_{REAL} + LL_{BDM}))$. LR follows a Chi-Square distribution with degrees of freedom equal the difference between the sum of the coefficients of the two unrestricted models and the sum of the coefficients of the restriced model. We mark the significance levels as follows: ***: $p < .001$: qchisq(1-0.01,2) = 9.210; **: $p < .05$: qchisq(1-0.05,2) = 5.991; *: $p < .1$: qchisq(1-0.10,2) = 4.605

Table F4

POINT ESTIMATES AND 95% CONFIDENCE INTERVALS FOR OPTIMAL PRICE, QUANTITY AND PROFITS*

| *Method* | *Optimal Price* | *Confidence Interval* | *Absolute difference to benchmark* | *Optimal Quantity* | *Confidence Interval* | *Absolute difference to benchmark* | *Optimal Profits* | *Confidence Interval* | *Absolute difference to benchmark* |
|---|---|---|---|---|---|---|---|---|---|
| OE | 343.983 | [313.452, 413.410] | 95.096 | .374 | [.230, .467] | .353 | 512,736.578 | [370,312.657, 589,808.561] | 503,873.4 |
| CBC | 336.898 | [300.243, 392.883] | 88.011 | .314 | [.203, .406] | .293 | 455,387.414 | [350,017.9, 524,268.868] | 446,524.2 |
| BDM | 248.887 | [214.349, 374.107] | n.a. | .021 | [.011, .032] | n.a. | 8,863.183 | [-3,028.554, 17,440.451] | n.a. |

Notes: Quantity scaled from [0,1], n.a. = not applicable
* The gray-shaded cells indicate that the confidence interval of the specific measure overlaps with the confidence interval of the corresponding benchmark measure obtained from our real purchase data. Hence, shaded areas imply no statistical difference between the estimated measure and the benchmark.

Table F5

MEAN TRANSPARANCY AND ACCEPTABILITY RATINGS PER EXPERIMENTAL TREATMENT GROUP

(STANDARD DEVIATIONS IN PARENTHESES)

| | *Method* | | |
|---|---|---|---|
| | *Study 2* | | |
| | *OE (n=152)* | *CBC (n=207)* | *BDM (n=94)* |
| This task was very easy to understand and complete. (1 = "not at all", 7 = "very much so") | $5.28^{ab}$ (1.605) | $5.60^{a}$ (1.417) | $5.68^{b}$ (1.428) |
| Is it clear to you why it is in your best interest to state exactly the price you are willing to pay for the cleaning product/ to choose exactly the product alternative that represents your true preferences as close as possible? (1 = "not at all", 7 = "very much so") | n.a. | n.a. | 6.13 (1.330) |
| Did you understand the buying process? (reverse scored: 1 = "very much so", 7 = "not at all") | n.a. | n.a. | 1.88 (1.335) |
| Did you perceive the buying process as fair? (1 = "not at all", 7 = "very much so") | n.a. | n.a. | 5.05 (1.792) |
| I will be happy to do this task again in the future. (1 = "not at all", 7 = "very much so") | 4.91 (1.806) | 5.06 (1.707) | 4.77 (2.086) |
| Incomplete surveys (in % of total online surveys started) | 23.25 | 21.29 | 32.89 |

[ab] Values with same superscripts differ (at $p < .10$ in a pairwise t-test)
Notes: n.a. = not applicable

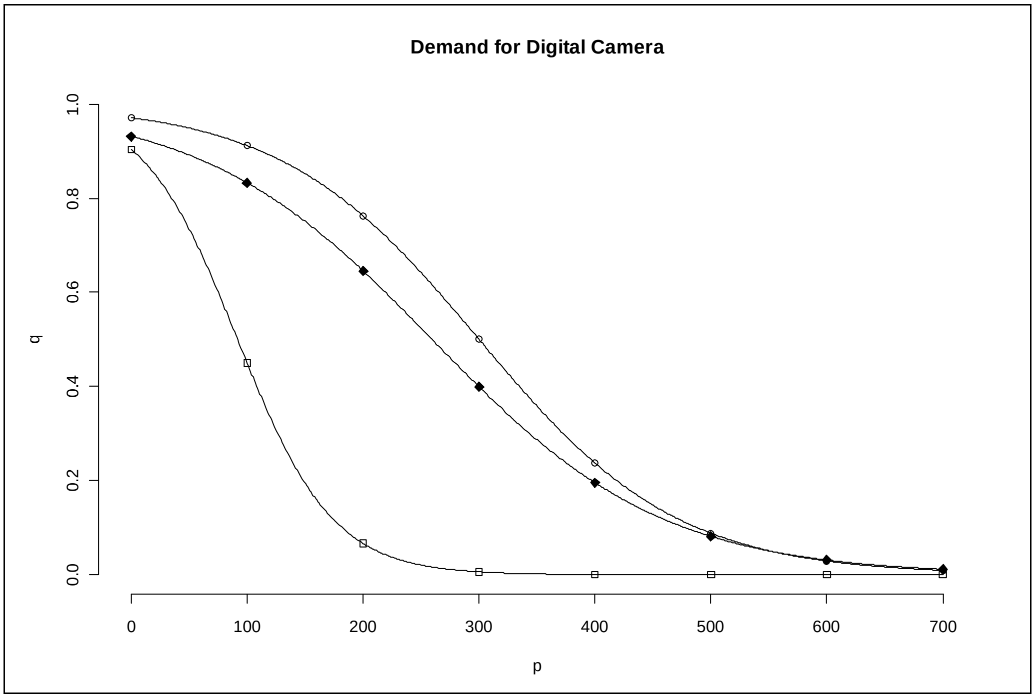
Table F6

LOGIT MODEL FITS TO OE, CBC, AND BDM SURVIVAL FUNCTIONS

| *Method* | *A* | *β* | *AIC* |
|---|---|---|---|
| OE | 3.500 | -.012 | 23.61 |
| CBC | 2.611 | -.010 | 126.90 |
| BDM | 2.236 | -.024 | 21.14 |

Figure F1

DEMAND CURVE FOR DIGITAL CAMERA




Notation: circles: OE, filled-diamonds: CBC, squares: BDM

## *REFERENCES WEB APPENDIX*

Backhaus, Klaus, Robert Wilken, Markus Voeth, and Christina Sichtmann (2005), "An Empirical Comparison of Methods to Measure Willingness to Pay by Examining the Hypothetical Bias," *International Journal of Market Research*, 47 (5), 543-62.

Blalock, Hubert M., Social Statistics, NY: McGraw-Hill, 1972: 406-7.

Dellaert, Benedict G., Aloys W. Borgers, and Harry J. Timmermans (1995), "A Day in the City, Using Conjoint Choice Experiments to Model Tourists Choice of Activity Packages," *Tourism Management*, 16, 347-53.

Ding, Min (2007), "An Incentive-Aligned Mechanism for Conjoint Analysis," *Journal of Marketing Research*, 44 (2), 214-23.

———, Rajdeep Grewal, and John Liechty (2005), "Incentive-Aligned Conjoint Analysis," *Journal of Marketing Research*, 42 (1), 67-82.

Efron, Bradley and Robert J. Tibshirani (1993), *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.

Haaijer, Rinus and Michael Wedel (2001), "Conjoint Choice Experiments: General Characteristics and Alternative Model Specifications," in *Conjoint Measurement – Methods and Applications*, Vol. 2, Anders Gustafsson, Andreas Herrmann, and Frank Huber, eds. Berlin, Springer Verlag, 319-60.

Hennig-Thurau, Thorsten, Victor Henning, Henrik Sattler, Felix Eggers, and Mark B. Houston (2007), “The Last Picture Show? Timing and Order of Movie Distribution Chanels,” *Journal of Marketing*, 71 (October), 63-83.

Hoogerbrugge, Marco and Kees van der Wagt (2006), “How Many Choice Tasks Should We Ask?,” *Sawtooth Software Conference Proceedings*, 97-109.

Jedidi, Kamel and Sharan Jagpal (2009), “Willingness to Pay: Measurement and Managerial Implications,” in *Handbook of Pricing Research in Marketing*, Vithala R. Rao, ed. Cheltenham: Edward Elgar Publishing, 37-60.

Kaas, Klaus P. and Heidrun Ruprecht (2006), “Are the Vickrey Auction and the BDM-Mechanism Really Incentive Compatible? Empirical Results and Optimal Bidding Strategies in the Case of Uncertain Willingness-to-Pay,” Schmalenbach Business Review, 55 (1), 37-55.

Kalish, Shlomo and Paul Nelson (1991), “A Comparison of Ranking, Rating and Reservation Price Measurement in Conjoint Analysis,” *Marketing Letters*, 2 (4), 327-35.

Keller, Gerald, Brian Warrack, and Henry Bartel (1990), *Statistics for Management and Economics: A Systematic Approach*, 2nd. ed., Belmont, CA: Wadsworth.

Lenk, Peter J., Wayne S. DeSarbo, Paul E. Green, and Martin R. Young (1996), “Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Desings,” *Marketing Science*, 15 (2), 173-91.

Lusk, Jayson L. and Ted C. Schroeder (2006) "Auction Bids and Shopping Choices," *Advances in Economic Analysis & Policy*, 6 (1), 1-37.

Moore, William L., Jason Gray-Lee, and Jordan J. Louviere (1998), “A Cross-Validity Comparison of Conjoint Analysis and Choice Models at Different Levels of Aggregation,” *Marketing Letters*, 9 (2), 195-207.

Oppewal, Harmen, Jordan J. Louviere, and Harry J. Timmermans (1994), “Modeling Hierarchical Conjoint Processes with Intregrated Choice Experiments,” *Journal of Marketing Research*, 31, 92-105.

Silva, Andres, Rodolfo M. Nayga Jr., Ben L. Campbell, and John Park, (2007), “On the Use of Valuation Mechanisms to Measure Consumers’ Willingness to Pay for Novel Products: A Comparison of Hypothetical and Non-Hypothetical Values,” *International Food and Agribusiness Management Review*, 10 (2), 165-80.

Sonnier, Garrett, Andrew Ainslie, and Thomas Otter (2007), “Heterogeneity Distributions of Willingness-to-Pay in Choice Models,” *Quantitative Marketing Economics*, 5 (3), 313–31.

Toubia, Olivier, Duncan I. Simester, John R. Hauser, and Ely Dahan (2003), “Fast Polyhedral Adaptive Conjoint Estimation,” *Marketing Science*, 22 (3), 273-303.

———, John R. Hauser, and Duncan I. Simester (2004), “Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis,“ *Journal of Marketing Research*, 61, February, 116-31.

Veisten, Knut (2007), “Willingness to Pay for Eco-labelled Wood Furniture: Choice-based Conjoint Analysis versus Open-ended Contingent Valuation,” *Journal of Forest Economics*, 13 (1), 29-48.

Voelckner, Franziska (2006), “An Empirical Comparison of Methods for Measuring Consumers’ Willingness to Pay,” *Marketing Letters*, 17 (2), 137-49.

Vriens, Marco, Harmen Oppewal, and Michel Wedel (1998), “Ratings-Based versus Choice-Based Latent Class Conjoint Models - An Empirical Comparison,” Journal of the Market Research Society, 40, 237-48.

Wertenbroch, Klaus and Bernd Skiera (2002), “Measuring Consumers’ Willingness to Pay at the Point of Purchase,” *Journal of Marketing Research*, 39 (2), 228-41.

Wittink, Dick R. and Trond Bergestuen (1999), "Forecasting with Conjoint Analysis," in *Principles of Forecasting: A Handbook for Researchers and Practitioners,* J. Scott Armstrong, ed. Norwell MA: Kluwer Academic, 147-67.