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

This study compares the performance of four commonly used approaches to measure consumers' willingness to pay with real purchase data (REAL): the open-ended (OE) question format; choice-based conjoint (CBC) analysis; Becker, DeGroot, and Marschak's (BDM) incentive-compatible mechanism; and incentive-aligned choice-based conjoint (ICBC) analysis. With this five-in-one approach, the authors test the relative strengths of the four measurement methods, using REAL as the benchmark, on the basis of statistical criteria and decision-relevant metrics. The results indicate that the BDM and ICBC approaches can pass statistical and decision-oriented tests. The authors find that respondents are more price sensitive in incentive-aligned settings than in non-incentive-aligned settings and the REAL setting. Furthermore, they find a large number of "none" choices under ICBC than under hypothetical conjoint analysis. This study uncovers an intriguing possibility: Even when the OE format and CBC analysis generate hypothetical bias, they may still lead to the right demand curves and right pricing decisions.

Keywords: market research, pricing, demand estimation, willingness to pay, hypothetical bias

How Should Consumers' Willingness to Pay Be Measured? An Empirical Comparison of State-of-the-Art Approaches

Accurately gauging consumers' willingness to pay (WTP)[1] for a product or service[2] is critical for formulating competitive strategies, conducting value audits, and developing new products (Anderson, Jain, and Chintagunta 1993). It is also important for implementing various pricing tactics, such as nonlinear pricing (Jedidi and Zhang 2002), one-to-one pricing (Shaffer and Zhang 2000), and targeted promotions (Shaffer and Zhang 1995). Not surprisingly, several approaches have been developed for this purpose (Cameron and James 1987; Jedidi and Zhang 2002). The

*Klaus M. Miller is Assistant Professor of Marketing (e-mail: miller@imu.unibe.ch), and Harley Krohmer is Professor of Marketing, Institute of Marketing and Management (e-mail: krohmer@imu.unibe.ch), University of Bern. Reto Hofstetter is Assistant Professor of Marketing, Institute of Information Management, University of St. Gallen (e-mail: reto.hofstetter@unisg.ch). Z. John Zhang is Murrel J. Ades Professor of Marketing, the Wharton School, University of Pennsylvania (e-mail: zjzhang@wharton.upenn.edu). This article is based on the first two authors' (cowritten) dissertation. The authors thank Sawtooth Software Inc. for generously sponsoring this research with the Sawtooth Software Grant and two anonymous companies for providing the stimuli used in this study. The authors further acknowledge helpful comments from Eric Bradlow, Christophe van den Bulte, Lutz Dümbgen, John Howell, Joffre Swait, and the two anonymous *JMR* reviewers. All remaining errors are the authors' responsibility. Ravi Dhar served as associate editor for this article.

[1]We take the standard economic view of consumer WTP and define it as the maximum price at or below which a consumer will definitely buy one unit of the product (Varian 1992). This corresponds to the concept of the floor reservation price that Wang, Venkatesh, and Chatterjee (2007) propose. However, we do not adopt their idea of conceptualizing WTP as a range. Instead, we view WTP as a point measure in line with and comparable to prior literature in economics on measuring consumer WTP (e.g., Wertenbroch and Skiera 2002). Furthermore, Wang, Venkatesh, and Chatterjee's procedure has been implemented so far only for direct measurements, not for indirect measurement approaches, such as conjoint analysis. Nevertheless, we account for the individual variation of consumer WTP by constructing appropriate confidence intervals for our WTP measures at the aggregate level (see the "Results" section).

[2]We focus on the WTP for the product as a whole, assuming known availability and awareness of the product. Our study does not address the WTP for features of a product.

primary distinctions among the approaches are whether they measure WTP directly or indirectly and whether they determine consumers' hypothetical WTP or actual WTP. In this article, we use a large-scale experimental design and field test to provide some preliminary evidence regarding the relative performance of four commonly used approaches to measure WTP (see Table 1).

In practice, some marketing researchers favor the direct approach, asking consumers directly to state their WTP for a specific product through, for example, an open-ended (OE) question format (Abrams 1964; Adam 1958; Arrow et al. 1993; Mitchell and Carson 1989). Others prefer an indirect approach, such as choice-based conjoint (CBC) analysis (Louviere and Woodworth 1983), in which WTP is calculated on the basis of consumers' choices among several product alternatives and a "none" choice option. However, neither method is foolproof. Many studies have shown that both direct and indirect approaches can generate inaccurate results for various psychological and technical reasons (Chernev 2003; Mitchell and Carson 1989; Orme 2003; Simon 1989; Steenkamp and Wittink 1994; Verlegh, Schifferstein, and Wittink 2002; Wittink, Krishnamurthi, and Nutter 1989). More fundamentally, both approaches measure consumers' hypothetical, rather than actual, WTP and thus can generate hypothetical bias, which the economics literature defines as the bias induced by the hypothetical nature of a task (Harrison and Rutström 2008).[3]

A direct approach to elicit actual WTP is a mechanism that Becker, DeGroot, and Marschak (1964; hereinafter, we refer to this as the BDM mechanism) propose, in which a participant is obligated to purchase a product if the price drawn from a lottery is less than or equal to his or her stated WTP (see also Wertenbroch and Skiera 2002). An indirect approach for determining actual WTP is the incentive-aligned choice-based conjoint (ICBC) analysis (Ding 2007), in which participants are also obligated to make a purchase based on WTP inferred from their revealed preference, using the BDM mechanism.[4] With more realistic economic incentives for survey respondents, these two approaches have generated good results in some applications (Ding 2007; Wertenbroch and Skiera 2002). However, an actual WTP generated with these methods may not always be accurate, because it may differ from the WTP shown in real consumer purchases (Harstad 2000; Kaas and Ruprecht 2006; Kagel, Harstad, and Levin 1987; Kagel and Levin 1993).

In this study, we develop a single-source data set. This enables us to compare the performance of these four approaches on the same task and relative to the same benchmark of real purchases to shed light on their strengths.

PRIOR RESEARCH

Many prior studies have examined the performance of various methods for determining WTP.[5] Some studies have shown an upward bias for a hypothetical OE question format as well as for hypothetical CBC analysis when comparing hypothetical WTP from these methods with actual WTP from BDM and ICBC (e.g., Lusk and Schroeder 2004; Wertenbroch and Skiera 2002). Other studies have gone one step further and used real purchase data as the benchmark (hereinafter, we refer to this benchmark as REAL) to assess hypothetical methods (OE and CBC) or incentive-aligned methods (BDM and ICBC; see Bohm, Lindén, and Sonnegård 1997; Frykblom 1997). The first attempt to compare all four methods (i.e., OE, BDM, CBC, and ICBC) with REAL in one study was that of Ding, Grewal, and Liechty (2005).[6] In two studies on inexpensive, frequently purchased products, they found that ICBC yielded the most valid results, followed by CBC, BDM, and OE, with respect to out-of-sample choice predictions. In contrast, we compare the same four methods with REAL in the context of measuring consumers' WTP on the basis of mean WTP, the resulting demand curves, and the method's ability to perform a certain pricing decision task.

In addition to comparing hypothetical and incentive-aligned methods to measure consumers' WTP, prior studies have compared direct or indirect approaches to measure WTP. Several studies confirm the existence of a significant difference between hypothetical direct approaches (OE) and hypothetical indirect approaches (CBC), based on their out-of-sample choice predictions and their resulting mean WTP and WTP distributions (Backhaus et al. 2005; Silva et al. 2007). Similarly, studies also confirm that incentive-aligned direct approaches (BDM) and incentive-aligned indirect approaches (ICBC) generate different results. However, although these studies document the differences, they do not investigate which method comes closer to consumers' real WTP.

In this article, we assess whether the four previously discussed approaches (OE, CBC, BDM, and ICBC) to measure consumers' WTP are statistically different from REAL and which of these approaches may lead to better pricing decisions.[7] This study contributes to pricing theory and practice in two ways. First, through a large-scale experimental

[3]In the literature on conjoint analysis, the term of the hypothetical bias is used differently, and the bias can result from any methods that use hypothetical profiles, which include non-incentive-aligned (hypothetical) and incentive-aligned conjoint analysis. In our study, we focus only on the bias in price but not in other partworths.

[4]Ding, Grewal, and Liechty (2005) and Dong, Ding, and Huber (2010) also propose a procedure for incentive-aligned conjoint analysis. Here, more than one product variation a participant is asked to evaluate is available for purchase. In most conjoint applications, however, conjoint practitioners and researchers have access only to one product variation. Therefore, we focus on Ding's (2007) proposed approach, which requires only one real product to be implemented.

[5]A comprehensive, detailed summary of previous research is available in Web Appendix A (http://www.marketingpower.com/jmrfeb11).

[6]Silva and colleagues (2007) also compare OE, BDM, CBC, and ICBC in a four-in-one study but do not provide a REAL benchmark, leaving the following question unasked: Which approach yields the most valid results?

[7]Park, Ding, and Rao (2008) propose a sequential incentive-compatible conjoint procedure for eliciting consumer WTP for attribute upgrades. We do not consider this method in our study because we focus on measuring consumer WTP for an entire product, not product attributes.

Table 1
ALTERNATIVE METHODS FOR MEASURING CONSUMERS' WTP

	Measurement	
Context	*Direct*	*Indirect*
Hypothetical WTP	OE question format	CBC analysis
Actual WTP	BDM mechanism	ICBC analysis

design and field test, we collect data from 1124 consumers and compare directly stated hypothetical WTP from an OE format, indirectly stated hypothetical WTP from CBC analysis, directly stated actual WTP under the BDM mechanism, indirectly stated actual WTP under ICBC, and REAL obtained from an online shop. This five-in-one study enables us to conduct a comprehensive assessment of each approach's ability to capture mean WTP and WTP distributions. Second, we make these comparisons on the basis of managerially relevant criteria, such as predicting the optimal price, quantity, and profits. These comparisons enable us to assess the suitability of these approaches as an aid to particular kinds of pricing decisions. They also help us better understand the strengths and weaknesses of each approach.

In the next section, we describe our data collection process. We then analyze our data and assess each of the four approaches. Finally, we offer some concluding remarks.

METHOD AND DATA COLLECTION

Participants and Stimulus

We collected data for the study through an online survey. To recruit participants, we sent 14,321 invitation e-mails to Swiss consumers. They included people from the student body (both undergraduate and graduate students and doctoral candidates) of a Swiss university, high school students, and other customers targeted by the company that supplied our stimulus. In addition, we distributed 5000 flyers in a Swiss city to invite other consumers to take part in the study. We motivated participation by offering a chance to win an MP3 player in a raffle.[8] The participants were further informed that their chance to win in the raffle was independent of their survey responses.[9] A total of 1124 people chose to participate. The stimulus we used was an innovative cleaning product for high-tech equipment that was not actually available in Switzerland at the time of the study and has no category competition.

Experimental Design

We developed five different independent experimental groups and used a between-subjects design. In the OE group, each participant directly stated his or her hypothetical WTP for the cleaning product. In the BDM group, we determined actual WTP by using the BDM mechanism, which we implemented in a way similar to Wertenbroch and Skiera (2002). Participants were told that they were obligated to buy the cleaning product at the randomly determined price if the price was less than or equal to their stated WTP. However, if the randomly determined price was higher, they would not be able to buy the product. This mechanism ensures that participants have no incentive to indicate a price that is higher or lower than their true WTP.

In the CBC group, we use a computer-generated CBC design (Louviere and Woodworth 1983). To administer the study, we used Sawtooth Software CBC/Web 6.4.2. We chose a CBC design because choice tasks are more immediate and concrete than rating or ranking tasks (Huber 1997). Moreover, CBC designs are frequently applied in pricing studies (e.g., Orme 2006). To determine appropriate attributes and attribute levels for our conjoint study, we conducted qualitative interviews with five consumers who were interested in buying the cleaning product and the product manager. We also conducted a pretest survey of 82 potential customers to identify five attributes associated with our cleaning product: (1) brand, (2) color, (3) durability, (4) cleaning power, and (5) price (see Table 2); this resulted in a $4^3 3^1 5^1$ attribute space. The product for our study was called CLEAN-A, a yellow cleaning compound with a durability of eight months, the cleaning power of 90%, and an intended retail price of 7.95 Swiss francs (CHF).[10]

In our study, we were careful to avoid the ad hoc nature of the price range effect (see Verlegh, Schifferstein, and Wittink 2002). We chose a realistic price range bounded by the high and low prices of comparable cleaning products commonly available in Switzerland. We also used the price coefficient relevant to the price range we were evaluating (Jedidi and Zhang 2002) whenever we needed to compute WTP. Furthermore, in our study, we estimated the net effect of the informational and allocative effect of price and made comparisons across different methods at this level.[11]

In the CBC group, we gave respondents seven choice sets and told them to imagine that they needed to choose among the product alternatives in an online shop "right here" and

[8]We used the MP3 player as a single incentive to motivate participation in our survey to recruit an adequate number of participants. However, the MP3 player was not connected to our stimulus and the incentive-aligned conditions under BDM and ICBC, in which proper incentives are offered to participants so that they are motivated to reveal their true preferences (for details, see Ding 2007; Wertenbroch and Skiera 2002).

[9]It is possible that some consumers took part in the survey just to win the MP3 player and were not interested in purchasing the cleaning product. However, there is no reason to believe that these participants affected our comparisons across different methods because they were randomly assigned to different task groups.

[10]All brand names are anonymized.

[11]Prior studies have shown that the two effects may be confounded in a preference model such that the estimated WTP measures may be negative (Jedidi and Jagpal 2009). To remedy this problem, Gautschi and Rao (1990), and more recently Voelckner (2008), propose a methodology to separate price effects in conjoint studies, which requires additional data collection. In our study, we followed Rao and Sattler's (2003) proposed approach to reduce the confounding problem by including the most salient product attributes to the consumers in the conjoint analysis, which we obtained in our pretest study. Furthermore, for our purpose, we do not need to decompose different price effects to arrive at WTP. In addition to our estimates being all inclusive, such price effects apply for the other methods studied, such as BDM, OE, and REAL. Because these effects affected all samples collected using the different methods in a similar way, we do not expect such effects to bias the results of our comparisons.

Table 2

ATTRIBUTES AND LEVELS INCLUDED IN CBC ANALYSIS

Attribute	*Levels*	*Number of Attribute Levels*
Brand	CLEAN-A, CLEAN-B, CLEAN-C, CLEAN-D	4
Color	Red, blue, green, yellow	4
Durability (period of usage)	2 months, 4 months, 6 months, 8 months	4
Cleaning power	Absorbs 90% of dust and dirt, absorbs 75% of dust and dirt, absorbs 60% of dust and dirt	3
Price	CHF1.59, CHF4.79, CHF7.95, CHF11.10, CHF14.30	5

"right now." Each choice set contained four cleaning products (i.e., conjoint stimuli) and a "none" purchase option. We described each conjoint stimulus by the five attributes obtained from pretesting. Attribute levels varied systematically (see Table 2). We created stimuli and conjoint choice sets according to a computer-generated randomized design that accounted for the design principles of minimal overlap, level balance, and orthogonality (Huber and Zwerina 1996).

In the ICBC group, the conjoint procedure was exactly the same as for the CBC group. The incentive-alignment mechanism we used was the same as that which Ding (2007) proposes. In our study, we informed participants that their responses in the conjoint task would be used to infer their WTP for a product and that after the completion of the survey, the product with the attributes preferred by the most people would be produced. The BDM mechanism embedded in the incentive-alignment mechanism ensures that participants have an incentive to reveal their true preferences in conjoint studies.

In the REAL group, we collected real transaction data by asking each participant whether he or she would be willing to buy the cleaning product at a certain price displayed in an online shop. The test site used for the experiment was similar to the real online shop of the cleaning product manufacturer. This elicitation of real WTP corresponds to a dichotomous choice question format often used in contingent valuation studies (see Brown et al. 1996). Real dichotomous choice data are incentive compatible and show high external validity because they are based on actual purchase observations made under realistic market conditions. Despite these favorable properties, the simulated Web-based store can suffer from measurement errors.[12] For the sake of comparability, price levels in the online shop corresponded to the price levels in our conjoint treatments (CBC and ICBC groups). We randomly assigned price levels to the participants, and each price level had an equal chance of appearing in the online shop.

To carry out the buying obligation under the incentive-aligned conditions (i.e., BDM and ICBC) and the REAL condition, we recorded the name and address of each participant in these experimental treatments. After the completion of the study, all participants who were obligated to buy (or agreed to buy the product in the online shop under the REAL condition) were sent the cleaning product, including an invoice, by mail. The invoice was due within 14 days and was payable with cash or a credit card. This payment process was officially approved by the proper authority of the university. Of all respondents in the BDM group, 93 (50.82%) were required to purchase the product, and in the ICBC group, 104 (68.87%) participants were required to purchase the product. In the REAL group, 107 consumers actually purchased the cleaning product. Only 4 respondents in the BDM condition and 3 respondents in the ICBC condition refused to comply with their purchase obligation. In the REAL condition, all customers paid for the product.

We gathered 279 responses in the OE group, 183 in the BDM group, 310 in the CBC group, 151 in the ICBC group, and 201 in the REAL group. The five experimental groups did not differ significantly in terms of sociodemographics or socioeconomic status. We performed a multivariate analysis of variance (Pillai-Spur: F = 1.361, $p = .112$) for age (F = 1.637, $p = .163$), sex (F = 1.014, $p = .399$), education (F = 1.716, $p = .144$), income (F = .850, $p = .494$), budget for cleaning products (F = 1.511, $p = .197$), and purchase interest (F = 1.632, $p = .164$).

Experimental Procedure

We divided our questionnaire into three parts. The first part described the product in the OE, BDM, and REAL groups as well as the relevant product attributes in the CBC and ICBC groups. The second part consisted of the WTP task in the different experimental treatment groups. In the third part of the questionnaire, we conducted a brief survey on sociodemographics and socioeconomics, and we made sure that the participants understood the WTP elicitation method to which they were exposed. We surveyed the respondents in the OE, CBC, BDM, and ICBC groups on their understanding of the different WTP measurement approaches.[13] We recorded several measures of transparency, acceptability, and completion rates (see Table 3).[14] In the REAL group, we surveyed participants only on their sociodemographics and economic background as part of the cleaning product's online buying process.

WTP Estimation Procedure

Figure 1 plots observed demand in each treatment, measured as the number of respondents whose WTP for the cleaning product is larger than a given p. To compare the different data sets, we used a common standard for comparison. We chose the survival functions of the form $q(p) = Pr(p \leq WTP)$ as the basis for all following analysis. Here, q(p) denotes the probability that a respondents' WTP is equal to or greater than a certain price, p. For the OE and BDM groups, we obtained respondents' hypothetical WTP and actual WTP directly from the survey data. To infer a participant's hypothetical WTP (actual WTP) from the preference data gathered in the CBC (ICBC) group, we followed Arora and Huber's (2001) approach, using the hierarchical Bayes (HB) routine to estimate individual-level partworths. Five of the seven choice tasks that we included in our survey

[12]There are several possible sources for errors. First, purchasing cleaning products online could be an unfamiliar distribution channel for some consumers. Second, the store allowed consumers to purchase only one product. Third, because the cleaning product was new to the market, no repeat purchases were observed. Fourth, the cleaning product was not displayed in a competitive setting; thus, consumers were unable to select this cleaning product from a group of competing products as they would be able to in a real store.

[13]Because respondents' understanding in the ICBC group was lower than that in the three other groups, we assessed the effect of respondent understanding on WTP in the ICBC group. To do so, we performed a median split based on respondent understanding and found that the mean WTP between respondents who found ICBC difficult to understand (n = 26) and those who understood it well (n = 125) showed no significant differences on a pairwise t-test (t = –.953, $p = .348$). This finding is in line with Ding's (2007, p. 222) notion that "some conjoint participants may not completely understand the mechanism but may believe that they will be better off if they try harder to fill it out as conscientiously (and as truthfully) as possible."

[14]The higher incompletion rate for incentive-aligned methods may be explained as follows: First, in contrast with Ding (2007), we did not prescreen respondents about their purchase interest. Second, incentive-aligned methods are inherently more complex than hypothetical methods. However, the methods are relatively more difficult but not so difficult that respondents cannot perform the tasks they are asked to perform.

Table 3
MEAN TRANSPARANCY AND ACCEPTABILITY RATINGS PER EXPERIMENTAL TREATMENT GROUP
(STANDARD DEVIATIONS IN PARENTHESES)

	Method: Study 1			
	OE (n = 279)	*BDM (n = 183)*	*CBC (n = 310)*	*ICBC (n = 151)*
This task was very easy to understand and complete. (1 = "not at all," 7 = "very much so")	5.49[a] (1.561)	5.72[b] (1.308)	6.01[a,b,c] (1.355)	5.06[a,b,c] (1.593)
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.	6.05[d] (1.244)	N.A.	5.73[d] (1.316)
Did you understand the buying process? (reverse scored: 1 = "very much so," 7 = "not at all")	N.A.	3.13 (2.168)	N.A.	3.13 (1.719)
Did you perceive the buying process as fair? (1 = "not at all," 7 = "very much so")	N.A.	5.33[e] (1.548)	N.A.	4.90[e] (1.441)
I will be happy to do this task again in the future. (1 = "not at all," 7 = "very much so")	5.00[f] (1.603)	4.93[f] (1.648)	5.59[f] (1.545)	5.07[f] (1.670)
Incomplete surveys (in % of total online surveys started)	24.43	36.78	22.67	46.59

[a, b, c, d, e, f]Values with the same superscripts differ at $p < .01$, $p < .02$, $p = .000$, $p = .021$, $p = .010$, and $p = .001$ in a pairwise t-test for a, b, c, d, e, and f, respectively.

Notes: N.A. = not applicable.

were randomly generated and used for partworth estimation.[15] We used the remaining two tasks for reliability and validity testing. We examined the sequence of output draws to ensure convergence had taken place. Because we were unable to observe any visible trends in the output draws after 100,000 burn-in iterations, we assumed that the data had converged. We then used the following 10,000 iterations for parameter estimation. We used individual-level partworth estimates to calculate individual consumers' maximum WTP.[16] Kohli and Mahajan (1991) propose the following relationship:

$$u_{it|\sim p} + v_i(p) \geq u_i^* + \in, \quad (1)$$

where $u_{it|\sim p}$ represents the total utility of the product excluding the utility of price to respondent i, $v_i(p)$ is the utility of a certain price level p, u_i^* represents the total utility of a certain threshold, and $\in$ stands for some positive number. Kohli and Mahajan (1991) use u_i^* as the status quo utility. To calculate consumers' WTP, we define u_i^* as the utility of not choosing the product (i.e., the utility of the "none" choice option in CBC/ICBC) so that Equation 1 is the same as Jedidi and Zhang's (2002 [see their Equation 1 on p. 1352]) definition.

More specifically, we use a piecewise linear approach to calculate individual WTP to mitigate the problem of fat tails that has hampered the use of conjoint analysis in estimating consumer WTP (see Web Appendix C at http://www.marketingpower.com/jmrfeb11).[17] The HB estimation yields estimates for $u_{it|\sim p}$ and u_i^* for each respondent i. For the purpose of the piecewise linear approach, we estimate $v_i(p)$ at each level of the price attribute p.[18] To find individual i's maximum WTP, we need to find the minimum $v_i(p)$ value

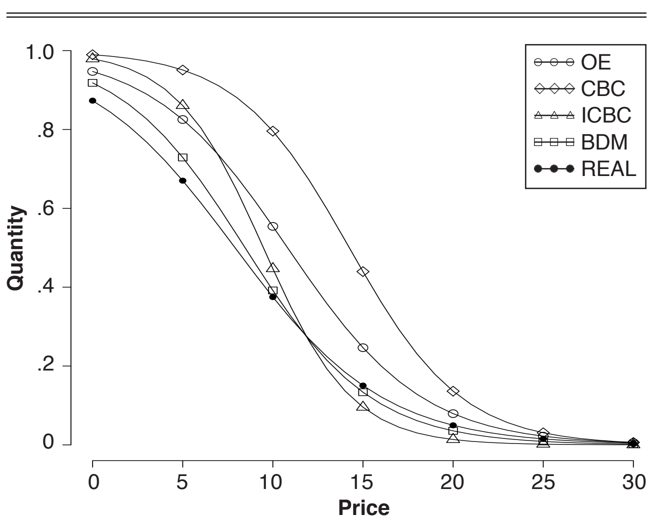
[15]We find that five choice tasks yield sufficiently valid partworth estimates for our case, as we show in Web Appendix B (http://www.marketingpower.com/jmrfeb11). However, because the number of choice sets in a conjoint analysis may influence parameter estimation, we do not know whether our results would hold for a larger number of choice sets per respondent, and further research can investigate this issue.

[16]The CBC/ICBC parameter estimates are confounded by the scale parameter of the Gumbel distribution (Swait and Louviere 1993). However, because any monotonic transformation of a utility function still represents the same preference structure, the scale parameter does not affect WTP estimates (i.e., the scale parameter drops out of WTP; see Sonnier, Ainslie, Otter 2007; Train 2003; Train and Weeks 2005; for a theoretical discussion, see also Jedidi and Zhang 2002).

[17]In the past, marketing researchers estimated a single linear coefficient for the price attribute and then calculated (marginal) WTP as a ratio of the coefficient of the nonprice attribute and the price attribute (for an overview, see Jedidi and Jagpal 2009). However, Sonnier, Ainslie, and Otter (2007) and Train and Weeks (2005) show that this traditional approach yields extremely fat tails in posterior WTP distributions, though it showed better in-sample fit of the data. They were able to mitigate this issue by reparameterizing the likelihood function such that it directly incorporates a measure for WTP. This makes it possible to control the prior WTP distribution in HB estimation. However, the studies differ for out-of-sample fit: While Train and Weeks find better out-of-sample fit for the traditional model, Sonnier, Ainslie, and Otter find the opposite.

[18]With Sawtooth's CBC/HB 4.6.4, the price attribute must be coded as "partworth" to obtain an estimate at each price level.

Figure 1
DEMAND CURVE FOR CLEANING PRODUCT



that still satisfies Equation 1. Using the inverse of $v_i(p)$, we can formulate WTP as follows:

$$\text{(2)} \qquad WTP = v_i^{-1}\left(u_i^* - u_{it|\sim p}\right).$$

In CBC analysis, the number of levels of the attribute price is limited. This only permits a discrete price level that is smaller or equal to maximum WTP. To find a more precise estimate for a consumer's maximum WTP, we must find a proper function $p = v_i^{-1}(u)$ that accounts for the discrete steps in price levels. In our case, we apply simple linear interpolation between the price levels for this purpose.[19]

We determined test–retest reliability of our conjoint data as an agreement between respondents' choices in the first and seventh choice tasks because the seventh task was identical to the first task (Ghiselli, Campbell, and Zedeck 1981). With identical choices by 88.06% in the CBC group and 87.42% in the ICBC group, reliability is good for the two subgroups.

For the REAL data, $q_{REAL}(p)$ represents the probability that a respondent answered "yes" to the dichotomous choice question at the price p. We used five price levels in the online shop, so we obtained five estimates for $q_{REAL}(p)$.

We determine face validity of WTP measures by correlating elicited WTP with the respondent's purchase interest. Face validity is high for all methods because correlations are significant (OE: r = .215, $p < .01$; CBC: r = .125, $p < .05$; BDM: r = .288, $p < .05$; ICBC: r = .346, $p < .01$).

To obtain parametric WTP estimates for the five experimental groups (see Figure 1), we first convert our WTP data to survival data. We then fit a logit model of the form $q(p) = Pr(buy|p) = \exp(\alpha + \beta \times p)/[1 + \exp(\alpha + \beta \times p)]$ (e.g., Balistreri et al. 2001; Bishop, Welsh, and Heberlein 1992; Brown et al. 1996; Wertenbroch and Skiera 2002), where p corresponds to the various price levels, α and β are the coefficients to be estimated, and Pr is a consumer's probability of purchasing the cleaning product at price level p. Maximization of the log-likelihood function yields estimates for the coefficients. Table 4 shows the parameter estimates.

[19]For a numerical example of this calculation, see Web Appendix D (http://www.marketingpower.com/jmrfeb11).

RESULTS

Analysis of Mean Bias

To analyze the data, we begin by using the common method of measuring the convergent validity of the elicited WTP by comparing mean hypothetical WTP with mean actual WTP. We constructed confidence intervals around the means and checked for overlaps (see Park, Loomis, and Creel 1991). We also performed pairwise t-tests whenever the data allowed us to do so. Table 5 shows the results of this analysis.

Our mean bias analysis uses the criterion of overlapping confidence intervals and cannot confirm the existence of a hypothetical bias. This result suggests that in our data set, all methods have a high convergent validity in measuring consumers' mean WTP. In the case of CBC, the confidence intervals only slightly overlap. Relative to the parametric REAL, CBC shows by far the largest bias ratio with 1.76, followed by OE with 1.30, ICBC with 1.11, and BDM with 1.06.

We also tested for mean differences among the different WTP measurement approaches using the t-test. Here, directly stated WTPs in BDM and OE differ significantly (Δ = CHF 2.06, t = 3.07, $p = .002$). Indirectly stated WTPs in CBC and ICBC also differ significantly (Δ = CHF 5.52, t = 9.69, $p < .001$). A reason for the much higher WTP estimates in CBC may be that there were many more "none" choices under ICBC.[20] In the ICBC group, 19% of the participants chose

[20]We thank an anonymous reviewer for this suggestion.

Table 4
LOGIT MODEL FITS TO OE, CBC, BDM, ICBC, AND REAL SURVIVAL FUNCTIONS

Method	α	β	*AIC*
OE	2.888	–.267	26.29
CBC	4.569	–.321	212.80
BDM	2.420	–.286	27.01
ICBC	3.857	–.407	98.55
REAL	1.929	–.244	7.94

Notes: AIC = Akaike information criterion.

Table 5
MEAN WTP, STANDARD ERRORS, AND 95% CONFIDENCE INTERVALS

Method	*n*	*Mean (Swiss Francs)*	*95% Confidence Interval*[d]	*Ratio of Hypothetical or Actual WTP to the Benchmark*
OE[a,c]	279	11.03	[10.196, 11.990]	1.30
CBC[b,c]	310	14.92	[14.164, 15.847][e]	1.76
BDM[a]	183	8.96	[8.097, 10.068]	1.06
ICBC[b]	151	9.39	[8.623, 10.134]	1.11
REAL (benchmark)	201	8.46	[5.348, 14.206]	N.A.

[a, b, c]Values with the same superscripts differ at $p < .05$, $p < .001$, and $p < .001$ in a pairwise t-test for superscripts a, b, and c, respectively.

[d]To calculate confidence intervals for REAL, we follow Park, Loomis, and Creel's (1991) approach, who apply Krinsky and Robb's (1986) procedure (10,000 iterations) to calculate confidence intervals for dichotomous choice WTP means based on Hanemann's (1984) model. For confidence intervals of the WTP means based on OE, BDM, CBC, and ICBC data, we apply a nonparametric approach because of the possibility of skewed WTP values. We used the bias-corrected and accelerated bootstrap percentile (10,000 iterations; see Efron and Tibshirani 1993) method for this purpose.

[e]Confidence intervals for estimates based on choice-based data can also be calculated through bootstrapping from the HB Markov chain Monte Carlo draws. These confidence intervals are much tighter than those calculated through bootstrapping from individual WTPs. For the sake of comparability, we used the latter approach because it can also be applied in the case of all other WTP measurement methods.

Notes: N.A. = not applicable.

the "none" choice option, and under hypothetical CBC, only 5% chose the "none" choice option; this difference results in a much larger intercept under CBC, but the difference in slopes is small. As a consequence, hypothetical CBC can be appropriate if a manager is interested mostly in the relative partworths of product attributes and price.[21] This finding is consistent with that of Ding, Grewal, and Liechty (2005) and indicates that under the hypothetical condition, participants behaved as if they were interested in purchasing one of the cleaning products, but they behaved differently when facing a real purchase decision under the incentive-aligned condition.

In summary, we find that mean WTP analysis shows statistically unbiased results for all methods when simply testing for overlap of confidence intervals. The more rigorous t-test shows that both hypothetical methods (i.e., OE and CBC) are significantly different from their incentive-aligned counterparts, indicating a hypothetical bias for product choice. The comparison of the hypothetical means with the real purchase benchmark shows that for our case of an inexpensive cleaning product, CBC is more biased in absolute terms than OE.

Analysis of Distribution Bias

Mean WTP is important for both value auditing and the valuation of a public good. However, for pricing decisions, even an accurate estimate of mean WTP may not be helpful to the marketing researcher in identifying optimal prices. For example, a unimodal distribution, compared with a bimodal distribution, even with a common mean, could imply a very different pricing structure. Therefore, we must consider the entire WTP distribution in assessing the performance of an approach, not just the mean (see Figure 1).

To compare WTP distributions, we first applied a nonparametric two-sided Kolmogorov–Smirnov (KS) test. We cannot apply this test to REAL because in this case, consumers' WTP is not available at an individual level. Given that incentive-aligned approaches (BDM and ICBC) provide good estimates for REAL, we use these data as the basis for this comparison. The KS test shows that the hypothetical WTP distributions of OE and CBC differ significantly from their incentive-aligned counterparts ($p < .001$; see Table 6) in ways that are consistent with our analysis of mean WTP values.

We also applied the likelihood ratio (LR) test, which enabled us to determine the overall fit of elicited demand curves to the true demand curve from the real data. In the OE, BDM, and ICBC group, the results suggest that we cannot reject the null hypothesis of equal distributions, which means that the demand elicited with these approaches does not differ significantly from actual demand. However, we find significant differences in WTP distributions between CBC data and REAL (LR statistic = 9.679, $p < .01$). The LR test also reveals that demand under BDM tracks real demand best, followed by ICBC, OE, and CBC. With this less stringent parametric test, even hypothetical methods can capture real demand well.

[21]A way to reduce or even eliminate hypothetical bias under CBC is to calibrate the data by correcting for the underselection of the "none" purchase option (as in so-called dual-response choice designs; see Brazell et al. 2006).

Table 6

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

Comparison	*Pearson's r*	*KS Test*	*LR Test[a]*
OE–REAL	.925	N.A.	1.819
CBC–REAL	.776	N.A	9.679**
BDM–REAL	.953	N.A.	.206
ICBC–REAL	.920	N.A.	2.309
OE–BDM	N.A.	.191**	2.957
CBC–ICBC	N.A.	.515**	84.560**
OE–CBC	N.A.	.471**	13.499**
BDM–ICBC	N.A.	.265**	4.765*

*$p < .1$.
**$p < .001$.
[a]The likelihood ratio (LR) test tests the null hypothesis of equal distributions. We calculated the LR as follows (e.g., for a comparison between REAL and BDM): $LR = -2 \times [LL_{pooled} - (LL_{REAL} + LL_{BDM})]$. The LR follows a chi-square distribution, with degrees of freedom equal to the difference between the sum of the coefficients of the two unrestricted models and the sum of the coefficients of the restricted model. We mark the significance levels as follows: *$p < .1$: $q\chi^2(1–.10, 2) = 4.605$; **$p < .05$: $q\chi^2(1–.05, 2) = 5.991$; and ***$p < .001$: $q\chi^2(1–.01, 2) = 9.210$.
Notes: N.A. = not applicable.

We also compared the fit of the form of the different hypothetical WTP and actual WTP demand curves from our survey data by correlating it with the nonparametric real WTP. All correlation coefficients have the expected (positive) signs and correlate significantly with real WTP.

Analysis of Outcome of a Business Decision

Do these differences matter for price setting or sales forecasting, the ultimate test of a successful approach? By examining how well each of these tools supports the business decision of choosing the profit-maximizing price, we can answer this question in our application.[22] We first compare the performance of the approaches in determining the demand curve within a range around the optimal price. We then examine the ability of the different approaches to explicitly forecast the optimal price, quantity, and profits.

Analysis of optimal price range. For the manufacturer of the cleaning product, the profit function is $\pi = ms \times q(p) \times (p - c)$, where q(p) is quantity scaled from [0, 1] given by $q(p) = [(e^{\alpha + \beta \times p})/(1 + e^{\alpha + \beta \times p})]$, ms is the market size, p is price, and c are the variable costs. Using the information from the manufacturer of the cleaning product, we know that c = CHF.85 and ms = 30,000, the expected sales in the target segment in Switzerland in 2008.[23] CLEAN-A did not have any direct competitors at the time of market introduction because it is a radically new way of cleaning high-tech equipment. However, after the product gains deeper penetration in the market, it may be necessary to consider indirect competitors (e.g., an air brush that can also clean high-

[22]In the conjoint analysis literature, it is becoming increasingly common to examine whether different methods lead to different managerial implications (e.g., by comparing optimal product line designs and pricing policies implied by various methods; see Belloni et al. 2008; Jedidi, Jagpal, and Manchanda 2003; Toubia, Hauser, and Garcia 2007).

[23]According to the manufacturer, variable costs did not depend on the number of units produced. However, variable costs may only be constant over a certain range around the actual quantity ordered by the manufacturer.

tech equipment).[24] Note that for the purpose of optimization, a nonzero fixed cost will not alter the outcome.

A statistical approach to construct a (KS) confidence range[25] around the profit maximizing price (p^* = CHF 8.50) is as follows[26]: We begin by constructing a confidence range around the parametric distribution of our real data.[27] The probability that real demand is in the confidence range is $1 - \alpha$. Thus, we can define $q_{lower}(p)$ as the lower bound of the confidence range, and $q_{upper}(p)$ as the upper bound. Then, the confidence range for the optimal profit is as follows:

$$\pi_{lower} = ms \times q_{lower}(p) \times (p - c), \text{ and}$$

$$\pi_{upper} = ms \times q_{upper}(p) \times (p - c).$$

Next, we construct a confidence range that contains the optimal price with a probability of $1 - \alpha$ by cutting with

$$\pi_{upper}^{-1}\left\{\max\left[\pi_{lower}(p)\right]\right\}.$$

This cut yields a confidence range of [CHF4.60, CHF17.39] at $\alpha = .10$ for the profit-maximizing price, given the demand and cost information we outlined previously (see Figure 2).[28]

We now determine whether the WTP distributions resulting from the various methods and their confidence range overlap with actual demand and the confidence range of the optimal price p^* for CLEAN-A. Our analysis shows (see Figure 3) that WTPs from BDM overlap at any given price point in the range of the profit-maximizing price. In other words, the BDM data are not statistically different from REAL. Furthermore, we find partial overlaps for OE and ICBC distributions in the range of the optimal price. However, CBC does not overlap at all in the relevant range for a pricing decision in our application. These findings should not be surprising, considering that in terms of the absolute distances between an elicited WTP distribution and the actual demand, BDM shows the least deviation from the benchmark ($\Delta = .170$), followed by ICBC ($\Delta = .661$), OE ($\Delta = 1.840$), and CBC ($\Delta = 4.376$).

Analysis of point estimates of optimal price, quantity, and profits. To assess the ability of the different approaches to correctly forecast optimal price, quantity, and profits, we use bias-corrected and accelerated bootstrapping (Efron and Tibshirani 1993) to construct 95% confidence intervals around these measures based on our real data. Using these confidence intervals as the benchmark, we compare them with the corresponding confidence intervals we construct using OE, BDM, CBC, and ICBC data. Table 7 shows the resulting point estimates and their confidence intervals.

For the optimal price, none of the estimates are statistically different from the benchmark. For the optimal quantity, only CBC yields a statistically different result. The point estimates for the optimal price and quantity from OE and CBC do not fall in the confidence intervals generated with other methods for the respective measures, though the confidence intervals overlap slightly. The findings for optimal profits paint a different picture. While forecasts from hypothetical approaches are statistically different from the benchmark, those from incentive-aligned approaches are not, though the absolute deviations from the benchmark are large for all approaches, especially for the hypothetical approaches. These findings suggest that in our application, incentive-aligned methods are better able to forecast not only optimal price and quantity but also profits.

However, what is most surprising from this analysis is the conclusion that the hypothetical methods can also do a good job in forecasting optimal price and quantity, even though they generate hypothetical bias. A possible explanation for this finding is that p^* in our case is less sensitive to changes in intercept (α) than to changes in slope (β).[29] When comparing intercepts and slopes, we find that hypothetical methods are largely biased in terms of intercept but not much in terms of slope (see Table 4). This implies that methods that yield demand curves, which are highly biased in terms of intercept but not so much in terms of slope, may still yield good estimates of optimal price p^*. By comparing the slopes (betas), we also observe that incentive-aligned methods yield steeper slopes than hypothetical methods and real data. This indicates that respondents are more price sensitive in our incentive-aligned settings. This is in line with Ding, Grewal, and Liechty (2005), Ding (2007 [see Experi-

[24]In this case, researchers would need to estimate demand not only for one product but also for all competitive products using all five approaches to gauge consumer WTP because this would allow for competitive equilibrium analysis.

[25]We call this a KS confidence range because we compare true distributions with empirical distributions for which the KS distribution is used for confidence.

[26]The profit-maximizing price is sensitive to parameterization and priors in HB estimation for WTP (see Sonnier, Ainslie, and Otter 2007). Because we did not explore this issue in our study, we do not know how our results would change if we had used different parameterizations and priors in HB estimation.

[27]$q_{lower}(p) = q(p) - \sqrt{-[\log(\alpha/2)/2]/n}$; $q_{upper}(p) = q(p) = q(p) + \sqrt{-[\log(\alpha/2)/2]/n}$; see Massart (1990).

[28]We were only able to obtain a min (π_{upper}) that is equal or lower than max (π_{lower}) for $\alpha = .1$, but not for $\alpha = .05$ or $\alpha = .01$.

[29]We thank an anonymous *JMR* reviewer for making this suggestion. Drawing the first derivative of the profit function, setting it to zero, and solving it for p^* yields the function for p^*. Marginal sensitivity of p^* to α and β can then be found by the partial derivations. Plugging in the values from the REAL distribution, we find that if α changes by 1%, p^* changes by .432%, but if β changes by 1%, p^* changes by .946%. This indicates that p^* is more than twice as sensitive to changes in β (slope) than to changes in α (intercept) (for these calculations, see Web Appendix E at http://www.marketingpower.com/jmrfeb11).

Figure 2

OPTIMAL PROFITS AND 90% CONFIDENCE RANGE FOR THE OPTIMAL PRICE

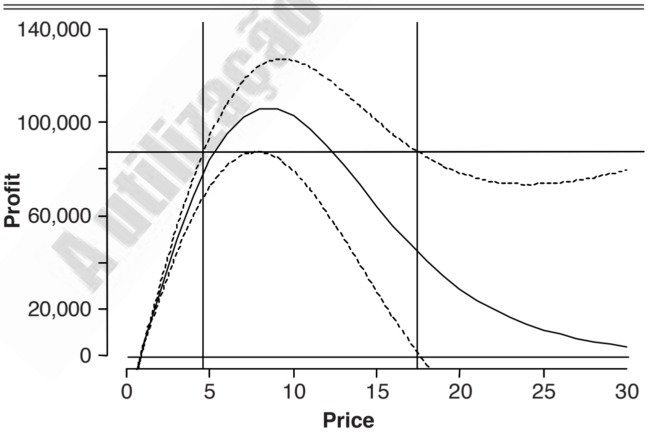
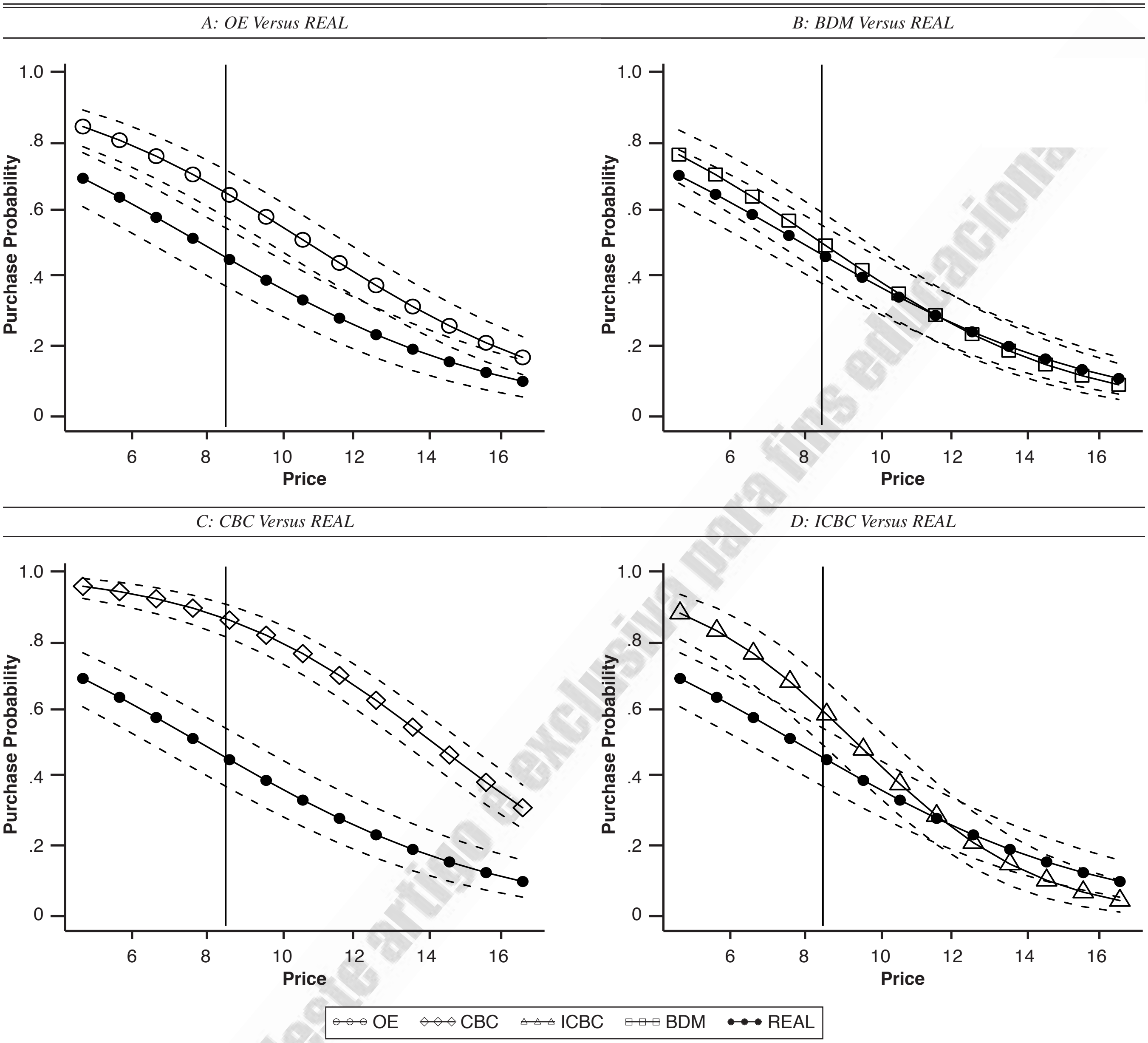


Figure 3

PLOTS OF THE PARAMETRIC WTP DISTRIBUTIONS IN THE OPTIMAL PRICE RANGE WITH 95% CONFIDENCE RANGES



ment 2]), and Wertenbroch and Skiera (2002), who find greater price sensitivity for incentive-aligned than for hypothetical approaches. Furthermore, in our data, we find that consumers in the REAL purchase task are the least price sensitive.

Here, we can also establish a rank order based on absolute deviations from the benchmark in our application: BDM yields the least deviation, followed by ICBC, OE, and CBC. The rank order is the same for all three point estimates (i.e., optimal price, quantity, and profits).

Our findings are significant in three ways. First, our analysis suggests that an incentive-aligned approach should be a more preferable choice for researchers and practitioners if it is not too costly to implement. However, applying incentive-aligned approaches may not always be feasible because of the availability of product prototypes or survey respondents and the legal restrictions on the types of marketing research that can be conducted. Second, the focus in previous research on hypothetical bias is perhaps beside the point for most marketing applications. Our analysis shows that even if a particular approach generates biased mean WTPs, and even if the estimated demand curve is different from the actual demand curve, the approach may still be useful in guiding marketing researchers to good pricing decisions. This result affirms the usefulness of OE and CBC, despite some concerns about the hypothetical nature of these approaches. Third, we find that OE can outperform CBC in estimating mean WTP and WTP distribution, as well as in making pricing decisions for an inexpensive, frequently purchased, nondurable product category.

Table 7

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	9.681	[8.408, 11.353]	1.181	.576	[.466, .659]	.112	152,483.7	[134,948.4, 163,644]	45,978.6
CBC	11.494	[10.352, 12.837]	2.994	.707	[.606, .777]	.243	225,799.3	[206,972.8, 237,568.7]	119,294.2
BDM	8.164	[6.782, 9.938]	.336	.522	[.393, .617]	.058	114,459.9	[95,014.62, 125,842.5]	7,954.8
ICBC	7.925	[6.896, 9.342]	.575	.652	[.523, .748]	.188	138,561.9	[120,520.8, 150,133.3]	32,056.8
REAL	8.500	[6.872, 12.299]	N.A.	.464	[.318, .602]	N.A.	106,505.1	[85,045.95, 123,140.7]	N.A.

Notes: Quantity scaled from [0, 1]. N.A. = not applicable. The 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 data. Thus, shaded areas imply no statistical difference between the estimated measure and the benchmark.

Note that the last finding stands in contrast with Ding, Grewal, and Liechty's (2005) finding that CBC yields better out-of-sample choice predictions than OE. One hypothesis is that CBC performs better when a product is less unique and faces more competing products, unlike our cleaning product.[30] Our analysis of a second data set from a digital compact camera offers some preliminary evidence in support of that hypothesis (see Web Appendix F at http://www.marketingpower.com/jmrfeb11). In that analysis, we find that both OE and CBC are significantly different from BDM, but CBC is less biased. We also find that in forecasting the optimal quantity and profits, neither OE nor CBC captures the true value, but CBC is marginally more accurate than OE in this application.

These findings are consistent with previous research on the strengths and weaknesses of different measurement approaches. Because consumer WTP is a context-sensitive construct (Thaler 1985), the suitability of a WTP measurement method can depend on how well such a method approximates the actual purchasing context of the underlying product and/or category. Thus, indirect approaches, such as conjoint analysis, may be better suited for the product category in which a more extensive decision process is involved (Backhaus et al. 2005; Voelckner 2006), while direct approaches are less suitable for infrequently purchased products (Gabor and Granger 1966) but more suitable for offerings without any explicit competitive offering (Huber 1997). Therefore, it is not surprising to observe in our application that direct methods are more suitable for relatively lower-priced, more frequently purchased, nondurable product categories with no direct competition, while indirect methods seem to be more suitable for relatively higher-priced, less frequently purchased product categories with significant competition (see preliminary evidence from our digital camera study in Web Appendix F [http://www.marketingpower.com/jmrfeb11]).

CONCLUSION

This study has several implications for applied marketing research. First, the results indicate that though the hypothetical approaches can generate mean WTP estimates that are not significantly different from actual WTP by some less restrictive statistical test, an incentive-aligned approach may be a better approach for the purpose of value auditing.

Second, we show that incentive-aligned methods can pass statistical and decision-oriented tests. Thus, we provide some preliminary evidence that marketing researchers would be well advised to use such methods, especially when a product prototype is available, the underlying product is not expensive, and the costs of fulfilling the actual buying obligations are low.

Third, the study uncovers an intriguing possibility: Although hypothetical methods are known to generate hypothetical bias, they may still lead to the right demand curves and right pricing decisions. In other words, an approach that generates a biased mean WTP value need not be dismissed entirely. Our analysis shows that a hypothetical approach, such as OE, has the potential to forecast an accurate demand curve. Furthermore, both hypothetical approaches we tested (i.e., OE and CBC) lead to pricing decisions that are not significantly different from the benchmark in a statistical sense (the differences are economically significant nonetheless). Researchers focusing on the mean accuracy of hypothetical approaches may have underestimated their value in guiding managerial decision making.

This study also confirms two empirical regularities about how measurement approaches interact with respondents as reported in previous research. First, in line with prior research, we find that incentive-aligned methods yield steeper demand slopes than hypothetical methods and real data. This indicates that respondents may be more price sensitive in incentive-aligned settings than in non-incentive-aligned settings and the real purchase setting. Second, consistent with Ding, Grewal, and Liechty's (2005) findings, we further find that one reason for the differences between non-incentive-aligned and incentive-aligned conjoint analysis may be the much larger number of "none" choices under ICBC. Under the hypothetical condition, participants behaved as if they were interested in purchasing one of the product offerings, but they acted differently when facing a real purchase decision under the incentive-aligned condition.

This study also has several limitations. First, in our CBC designs, we used five tasks for parameter estimation. Although this seems adequate for our case, using five tasks is on the lower side of practice. Second, we assess the relative performance of various approaches to measure consumers' WTP in forecasting the optimal price, quantity, and profits assuming CLEAN-A had a monopoly in high-tech cleaning products. Further research could extend our analysis to a competitive situation. Third, our comparisons are limited to four commonly used approaches to measure consumer WTP. Therefore, future studies could extend our analysis to other approaches that measure WTP, such as dual response choice designs (see Brazell et al. 2006) and the upgrading method (Park, Ding, and Rao 2008). Fourth, the statistical test we used to evaluate the outcome of the business decision (nonoverlapping confidence intervals) is not that stringent (Poe, Severance-Lossin, and Welsh 1994). Our results might change if a more rigorous statistical test is applied. A final limitation is the low completion rate and the low understanding under incentive-aligned settings. Future studies could try to reproduce our results by prescreening respondents.

The results from the study also raise three questions for further research. First, would these results hold true for other product categories? If our conclusions carry over to many other product categories, this could lend support to marketing researchers who favor simple, hypothetical approaches. The reproduction of our results using different conjoint stimuli, a different number of attribute levels, and different purchasing contexts is needed before we can confidently answer this question. Second, would incentive-aligned approaches (i.e., BDM and ICBC) yield unbiased results for other product categories? It would be worthwhile to discover whether hypothetical approaches can actually outperform incentive-aligned approaches, for example, in more expensive, infrequently purchased product categories in which strategic underbidding may be more likely and incentive alignment may not be as effective as in less expensive product categories. Third, what are the conditions under which CBC is the best choice? Additional studies are

[30]We thank an anonymous reviewer for this suggestion.

needed to delineate such conditions and to improve the approach. We used a piecewise linear interpolation to derive consumer WTP. Future studies could perform the same comparisons using WTP approaches that Sonnier, Ainslie, and Otter (2007) and Train and Weeks (2005) propose. The answers to these three questions should help clarify the best methods for generating an accurate WTP in an even wider range of circumstances.

REFERENCES

Abrams, Jack (1964), "A New Method for Testing Pricing Decisions," *Journal of Marketing*, 28 (January), 6–9.

Adam, Daniel (1958), *Les Réactions du Consommateur Devant le Prix*. Paris: Sedes.

Anderson, James C., Dipak Jain, and Pradeep K. Chintagunta (1993), "Understanding Customer Value in Business Markets: Methods of Customer Value Assessment," *Journal of Business-to-Business Marketing*, 1 (1), 3–30.

Arora, Neeraj and Joel Huber (2001), "Improving Parameter Estimates and Model Prediction by Aggregate Customization in Choice Experiments," *Journal of Consumer Research*, 28 (September), 273–83.

Arrow, Kenneth, Robert Solow, Paul R. Portnoy, Edward E. Leamer, Roy Radner, and Howard Schuman (1993), "Report from the NOAA Panel on Contingent Valuation," *Federal Register*, 58 (10), 4601–4614.

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.

Balistreri, Edward, Gary McClelland, Gregory Poe, and William Schulze (2001), "Can Hypothetical Questions Reveal True Values? A Laboratory Comparison of Dichotomous Choice and Open-Ended Contingent Values with Auction Values," *Environmental and Resource Economics*, 18 (3), 275–92.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marschak (1964), "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, 9 (3), 226–32.

Belloni, Alexandre, Robert Freund, Matthew Selove, and Duncan Simester (2008), "Optimizing Product Line Designs: Efficient Methods and Comparisons," *Management Science*, 54 (9), 1544–52.

Bishop, Richard C., Michael P. Welsh, and Thomas A. Heberlein (1992), "Some Experimental Evidence on the Validity of Contingent Valuation," working paper, Department of Agricultural Economics, University of Wisconsin.

Bohm, Peter, Johan Lindén, and Joakim Sonnegård (1997), "Eliciting Reservation Prices: Becker-DeGroot-Marschak Mechanisms vs. Markets," *The Economic Journal*, 107 (443), 1079–1089.

Brazell, Jeff D., Christopher G. Diener, Ekaterina Karniouchina, William L. Moore, Valerie Severin, and Pierre-Francois Uldry (2006), "The No-Choice Option and Dual Response Choice Designs," *Marketing Letters*, 17 (4), 255–68.

Brown, Thomas C., Patricia A. Champ, Richard C. Bishop, and Daniel W. McCollum (1996), "Which Response Format Reveals the Truth about Donations to a Public Good," *Land Economics*, 72 (2), 152–66.

Cameron, Trudy A. and Michelle D. James (1987), "Estimating Willingness to Pay from Survey Data: An Alternative to Pretest-Market Evaluation Procedure," *Journal of Marketing Research*, 26 (November), 389–95.

Chernev, Alexander (2003), "Reverse Pricing and Online Price Elicitation Strategies in Consumer Choice," *Journal of Consumer Psychology*, 13 (1–2), 51–62.

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

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

Dong, Songting, Min Ding, and Joel Huber (2010), "A Simple Mechanism to Incentive-Align Conjoint Experiments," *Internationl Journal of Research in Marketing*, 24 (4), 312–23.

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

Frykblom, Peter (1997), "Hypothetical Question Modes and Real Willingness to Pay," *Journal of Environmental Economics and Management*, 34 (3), 275–87.

Gabor, André and Clive W.J. Granger (1966), "Price as an Indicator of Quality: Report on an Enquiry," *Economica*, 33 (February), 43–70.

Gautschi, David A. and Vithala R. Rao (1990), "A Methodology for Specification and Aggregation in Product Concept Testing," in *Telecommunications Demand Modelling: An Integrated View*, Alain de Fontenay, Mary H. Shugard, and David S. Sibley, eds. Amsterdam: North-Holland, 37–63.

Ghiselli, Edwin E., John P. Campbell, and Sheldon Zedeck (1981), *Measurement Theory for the Behavioral Sciences*. San Francisco: W.H. Freeman.

Hanemann, W. Michael (1984), "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses," *American Journal of Agricultural Economics*, 66 (3), 332–41.

Harrison, Glenn W. and E. Elisabet Rutström (2008), "Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods," in *Handbook of Experimental Economics Results*, Charles R. Plott and Vernon L. Smith, eds. Amsterdam: North-Holland, 727–67.

Harstad, Ronald M. (2000), "Dominant Strategy Adoption and Bidders' Experience with Pricing Rules," *Experimental Economics*, 3 (3), 261–80.

Huber, Joel (1997), "What We Have Learned from 20 Years of Conjoint Research: When to Use Self-Explicated, Graded Pairs, Full Profiles, or Choice Experiments," in *Sawtooth Software Conference Proceedings*. Sequim, WA: Sawtooth Software, [available at http://www.sawtoothsoftware.com/download/techpap/1997Proceedings.pdf#page=257].

——— and Klaus Zwerina (1996), "The Importance of Utility Balance in Efficient Choice Designs," *Journal of Marketing Research*, 33 (August), 307–317.

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, UK: Edward Elgar Publishing, 37–60.

———, ———, and Puneet Manchanda (2003), "Measuring Heterogeneous Reservation Prices for Product Bundles," *Marketing Science*, 22 (2), 107–130.

——— and Z. John Zhang (2002), "Augmenting Conjoint Analysis to Estimate Consumer Reservation Price," *Management Science*, 48 (10), 1350–1368.

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.

Kagel, John H., Ronald M. Harstad, and Dan Levin (1987), "Information Impact and Allocation Rules in Auctions with Affiliated Private Values: A Laboratory Study," *Econometrica*, 55 (6), 1275–1304.

——— and Dan Levin (1993), "Independent Private Value Auctions: Bidder Behaviour in First-, Second- and Third-Price Auctions with Varying Numbers of Bidders," *Economic Journal*, 103 (419), 868–79.

Kohli, Rajeev and Vijay Mahajan (1991), "A Reservation-Price Model for Optimal Pricing of Multiattribute Products in Conjoint Analysis," *Journal of Marketing Research*, 28 (August), 347–54.

Krinsky, Itzhak and A. Leslie Robb (1986), "On Approximating the Statistical Properties of Elasticities," *Review of Economics and Statistics*, 68 (November), 715–19.

Louviere, Jordan J. and George Woodworth (1983), "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data," *Journal of Marketing Research*, 20 (November), 350–67.

Lusk, Jayson L. and Ted C. Schroeder (2004), "Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks," *American Journal of Agricultural Economics*, 86 (2), 467–82.

Massart, Pascal (1990), "The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality," *Annals of Probability*, 18 (3), 1269–83.

Mitchell, Robert C. and Richard T. Carson (1989), *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington, DC: Resources for the Future.

Orme, Bryan K. (2003), "Which Conjoint Method Should I Use?" Research Paper Series, Sawtooth Software.

——— (2006), *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*. Madison: Research Publishers.

Park, Timothy, John B. Loomis, and Michael Creel (1991), "Confidence Intervals for Evaluating Benefits Estimates from Dichotomous Choice Contingent Valuation Studies," *Land Economics*, 67 (1), 64–73.

Park, Young-Hoon, Min Ding, and Vithala R. Rao (2008), "Eliciting Preference for Complex Products: A Web-Based Eliciting Upgrading Method," *Journal of Marketing Research*, 45 (October), 562–74.

Poe, Gregory L., Eric K. Severance-Lossin, and Mike P. Welsh (1994), "Measuring the Difference (X-Y) of Simulated Distributions: A Convolutions Approach," *American Journal of Agricultural Economics*, 76 (November), 904–915.

Rao, Vithala R. and Henrik Sattler (2003), "Measurement of Price Effects with Conjoint Analysis: Separating Informational and Allocative Effects of Price," in *Conjoint Measurement. Methods and Applications*, Anders Gustaffsson, Andreas Herrmann, and Frank Huber, eds. Berlin: Springer, 47–66.

Shaffer, Greg and Z. John Zhang (1995), "Competitive Coupon Targeting," *Marketing Science*, 14 (4), 395–416.

——— and ——— (2000), "Competitive Pay to Switch or Pay Not to Switch: Third Degree Price Discrimination in Markets with Switching Costs," *Journal of Economics & Management Strategy*, 9 (3), 397–424.

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.

Simon, Hermann (1989), *Price Management*. Amsterdam: Elsevier Science Publishers.

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

Steenkamp, Jan-Benedict E.M. and Dick R. Wittink (1994), "The Metric Quality of Full-Profile Judgments and the Number-of-Attribute Levels Effect in Conjoint Analysis," *International Journal of Research in Marketing*, 11 (3), 275–86.

Swait, Joffre and Jordan J. Louviere (1993), "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models," *Journal of Marketing Research*, 30 (August), 305–314.

Thaler, Richard (1985), "Mental Accounting and Consumer Choice," *Marketing Science*, 4 (3), 199–214.

Toubia, Olivier, John Hauser, and Rosanna Garcia (2007), "Probabilistic Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis: Theory and Application," *Marketing Science*, 26 (5), 596–610.

Train, Kenneth (2003), *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press.

——— and Melvyn Weeks (2005), "Discrete Choice Models in Preference Space and Willingness-to-Pay Space," in *Applications of Simulation Methods in Environmental and Resource Economics*, Riccardo Scarpa and Anna A. Alberini, eds. Dordrecht, the Netherlands: Springer, 1–16.

Varian, Hal R. (1992), *Microeconomic Analysis*, Vol. 3. New York: W.W. Norton.

Verlegh, Peeter W.J., Hendrik N.J. Schifferstein, and Dick R. Wittink (2002), "Range and Number-of-Levels Effects in Derived and Stated Measures of Attribute Importance," *Marketing Letters*, 13 (1), 41–52.

Voelckner, Franziska (2006), "An Empirical Comparison of Methods for Measuring Consumers' Willingness to Pay," *Marketing Letters*, 17 (2), 137–49.

——— (2008), "The Dual Role of Price: Decomposing Consumers' Reactions to Price," *Journal of the Academy of Marketing Science*, 36 (3), 359–77.

Wang, Tuo, R. Venkatesh, and Rabikar Chatterjee (2007), "Reservation Price as a Range: An Incentive-Compatible Measurement Approach," *Journal of Marketing Research*, 44 (May), 200–213.

Wertenbroch, Klaus and Bernd Skiera (2002), "Measuring Consumers' Willingness to Pay at the Point of Purchase," *Journal of Marketing Research*, 39 (May), 228–41.

Wittink, Dick A., Lakshman Krishnamurthi, and Julia B. Nutter (1989), "The Effect of Differences in the Number of Attribute Levels on Conjoint Results," *Marketing Letters*, 1 (2), 113–23.

Fonte: Journal of Marketing Research (JMR), v. 48, n. 1, p. 172-184, 2011. [Base de Dados]. Disponível em: <http://web.ebscohost.com>. Acesso em: 2 mar. 2011.