Issues in the Use of Ratings-based Versus Choice-based Conjoint Analysis in Operations Management Research

Ekaterina V. Karniouchina[1]

William L. Moore[2]

Bo van der Rhee[3]

Rohit Verma[4]

Conjoint analysis has played an important role in helping make a number of operations management decisions including product and service design, supplier selection, and service operations capacity. Many recent advances in this area have raised questions about the most appropriate form of conjoint analysis for this research. We review recent developments in the literature and provide new evidence on how the choice between ratings- and choice-based conjoint models might affect the estimates of customer demand used in operations management models.

The biggest systematic difference between ratings-based (RB) and choice-based (CB) parameters is consistent with the compatibility effect, i.e., some enriched attributes like brand name tend to be more important in RB models and some comparable attributes like price are likely to be more important in CB models. Still, there were reasonably small differences between choice- and ratings-based parameters. Parameter similarity was also seen in the lack of differences both in the choice share validations when the ''keep on shopping'' alternative was not considered and in the profiles that were predicted to maximize choice shares. This suggests that the two approaches will produce similar estimates of the relative importance of various attributes.

In spite of demonstrated success with each method, several reasons lead us to recommend the use of hierarchical Bayesian choice-based conjoint models. First, the slightly higher individual hit rate validations give us greater confidence in predictive accuracy overall as well as an increased ability to target individual customers. Additionally, the greater ease of modeling both changes in market size and competitive reactions are attractive benefits of choice-based models.

[1] Chapman University, Orange, CA, USA
[2] David Eccles School of Business, University of Utah, Salt Lake City, UT, USA
[3] Cluster Marketing & Supply Chain Management, Nyenrode Business University, Breukelen, The Netherlands
[4] School of Hotel Administration, Cornell University, 338 Statler Hall, Ithaca, NY 14853, USA

Introduction

Many critical business decisions reside at the interface between marketing and operations as these two functional areas attempt to resolve the inherent tension between product differentiation and operational complexity (Krishnan and Ulrich, 2001). Cross functional marketing-operations decisions include, but are not limited to, product design (we use the word ''product'' to refer to products and services, as well as their combinations), product line and portfolio optimization, capacity planning, customer support management, as well as volume and mix flexibility decisions. As the economy is gaining a greater service focus, these integrated decisions are becoming even more critical as they are most imperative in dynamic service environments (Karmarkar, 1996). As a result in many companies, a focus on processes, cross-functional teams, and the need for speed are eroding traditional functional boundaries. Similarly, academics (e.g., Karmarkar, 1996; Krishnan and Ulrich, 2001; Schmidt and Porteus, 2000; Tatikonda and Montoya-Weiss, 2001) recognize the need for a broader problem focus. One of the challenges of cross functional research is that in many cases it needs to meet the publication standards of several base disciplines, which requires one to keep up with the literature in multiple subject areas.

One of these areas of increasing cross-disciplinary interest is the use of conjoint analysis in operations management research. Conjoint analysis has been used in various business applications for almost 40 years. Many applications of conjoint analysis are of great relevance to operations management scholars. Conjoint analysis has been applied to inform product line decisions (Yano and Dobson, 1998) with an emphasis on platforms (Moore et al., 1999); shared costs (Morgan et al., 2001); engineering constraints (Michalek et al., 2005); and robust product design (Luo et al., 2005). This research method also shed light on issues related to the timing of the next generation of high technology products (Kim and Srinivasan, 2006). Additionally, conjoint analysis has been used to assess tradeoffs and switching inertia associated with supplier selection (Verma and Pullman, 1998; Li et al., 2006). Other applications have focused on optimal service design considering operational difficulty (Verma et al., 2001) and operations capacity (Pullman and Moore, 1999; Goodale et al., 2003). Verma et al. (2006, 2008), Ding et al. (2007) and Victorino et al. (2005) have applied conjoint models in not-for-profit, financial, and hospitality service settings. Furthermore, an emerging emphasis on incorporating behavioral aspects into manufacturing and service operations models (Bendoly et al., 2006), portends future growth of conjoint analysis and related approaches in the operations management literature. The importance of conjoint analysis to readers the European Journal of Operational Research can be seen by a series of articles applying conjoint analysis to such topics as e-commerce (Butler et al., 2008), banking (Mankila, 2004), sales promotion design (Nair and Tarasewich, 2003), urban quality of life (Ülengin et al., 2001), forecasting (Natter and Feurstein, 2002; Turksen and Willson, 1995), product and product line design (Scholl et al., 2005; Tarasewich and McMullen, 2001), and pricing (Day and Venkataramanan, 2006; Green and Krieger, 1992).

Therefore, our purpose is to review recent developments and provide new evidence on how the choice of different variants of conjoint analysis might affect study results. In particular, we examine differences between ratings- and choice-based conjoint analyses. Our intended audience includes operations management researchers who want to use conjoint analysis in their papers to measure customer utility or demand as part of their models as well as marketers interested in performing conjoint analysis.

For many years, ratings-based (RB) conjoint analysis (where respondents rate their preference for different product profiles and individual-level attribute partworths are estimated with OLS regression) was more popular than choice-based (CB) conjoint analysis (where respondents make choices from a series of sets of product profiles and more aggregate-level parameters are estimated with logit or probit models). In the mid 1990's hierarchical Bayesian (HB) methods were introduced to both RB and CB conjoint analyses, which allowed more accurate estimates of individual- level parameters with fewer observations (e.g., Allenby and Lenk, 1994; Lenk et al., 1996). While HB methods benefited both RB and CB approaches, the impact on CB methods was greater as one was now able to estimate individual-level preferences, where only segment- and aggregate-level estimates had been available in the past. This new capability has shifted the preference of some academics and consultants to CB methods. The existence of two methods, both of which are able to capture individual-level preferences has led to questions about systematic differences in results that one would get when using these different approaches and which method should be used when attempting to design new and improved products and services.

Researchers have long recognized several fundamental differences between RB and CB methodologies. For instance, greater task similarity between choices and market behavior may lead to greater external validity (Elrod et al., 1992). Still, RB methods may be just as successful due to their ability to mirror marketplace simplification (Huber, 1987). An individual choice task provides less information than a ratings task but it may be easier for respondents. While CB conjoint models produce choice probabilities directly, RB models use ad hoc rules, like maximum utility or share of preference, to convert preference ratings into choice probabilities. Even though these rules have long proven successful in the marketing literature (e.g., Green et al., 1981) they can be seen as arbitrary. Operationally, RB studies are easier to design and estimate. Finally, the fact that preferences are not invariant across elicitation processes or contexts (e.g., Payne et al., 1992) may lead to systematic differences in the parameters estimated with RB and CB conjoint analyses.

The remainder of this paper is organized as follows: first we present the background related to differences between ratings and choice-based models and outline some theoretical reasons why the parameter estimates obtained from these two classes of models may differ systematically. Next we provide an empirical test where several models are estimated based on choice- vs. judgment-based tasks and examine the differences in resulting coefficients and corresponding model fits. Finally, we summarize the implications of our research as well as

provide managerial guidance regarding the choice between different estimation techniques (OLS, Hierarchical Bayes, and latent segment) as well as choice elicitation methods (choice-based versus ratings-based tasks).

Background

Judgment versus Choice

As their names imply, a fundamental difference between ratings-and choice-based conjoint analyses is that the former uses judgments (i.e., preference ratings) to elicit preferences while that latter uses choices. Behavioral decision theorists have long compared preferences elicited in different manners. Normative models treat judgment and choice equivalently supporting the common sense assumption that one should expect choices and judgments to possess descriptive and procedural invariance. In conjoint analysis this means that if a person rated one profile higher than another on a likelihood of purchase scale, that person would also choose the first profile over the second in a choice task. However, a large body of empirical evidence contradicts both descriptive and procedural invariance (e.g., Payne et al., 1992). The common assumption that preferences are often constructed rather than revealed (e.g., Bettman et al., 1998) applies to conjoint tasks. While subjects are familiar with the underlying attributes and levels, preferences for specific profiles need to be constructed. When preferences are constructed, the information and strategies are contingent on tasks, contexts, and individual differences.

Different response modes, such as judgment and choice can evoke different strategies. People may have ''meta-goals" for each decision such as maximizing accuracy or justifiability or minimizing effort, regret, or conflict (Bettman et al., 1998; Payne et al., 1992). The simplest or best process for deciding which of several alternatives is most preferred is not necessarily the simplest or best process for determining one's preference for each of the alternatives individually.

Tradeoffs between accuracy and effort suggest that in complex situations respondents tend to focus on more important pieces of information – either attributes (prominence effect) or attribute levels (level focusing effect) – or on information that is easier to translate into a decision (compatibility effect). If these simplification strategies vary persistently across tasks, violation of procedural and descriptive invariance would lead to systematic variation of the parameter estimates obtained from choice- and ratings-based tasks. We now discuss these three effects in more detail.

Prominence Effect

Fischer et al. (1999) call the prominence effect, which is the tendency to give more weight to the more important attributes in one preference elicitation task than another, one of the strongest preference biases documented in the decision making literature. Fischer and Hawkins (1993) found that the prominence effect was greater for choices than ratings.

However, others (e.g., Goldstein and Einhorn, 1987; Tversky et al., 1988) have found the opposite.

Scale Compatibility

A number of researchers have examined scale compatibility under the assumption that information requiring the smallest amount of transformation will be attended to more. For example, Hsee (1996) and Nowlis and Simonson (1997) have studied whether an attribute's compatibility with the preference elicitation task influences its importance. They found that comparable attributes, i.e., ones that produce clear and unambiguous comparisons, like price, were more important in comparison-based tasks. Conversely, enriched attributes that are less comparable, but are more easily evaluated on their own, like brand name, were more important in separate evaluations of individual options. This research suggests that enriched (comparable) attributes would be relatively more important in ratings-based (choice-based) conjoint tasks.

Level Focusing/Loss Aversion

A level focusing effect is observed when subjects focus on certain levels of an attribute more than others. A common example is to exhibit a tendency for diminishing returns to improvement or equivalently, a loss aversion for deterioration in an attribute (Kahneman and Tversky, 1979). Across three experimental conditions, Weitz and Wright (1979) found that adding cut-off variables to a linear model of product ratings increased explanatory power significantly in 27–60% of the cases. Orme et al. (1997) found a greater loss aversion in choice tasks compared to judgment tasks.

Summary

These differences raise questions about the similarity of the estimated parameters and choice predictions resulting from choice and ratings tasks. Therefore, we will examine whether there are systematic differences in the parameters estimated by RB and CB conjoint models.

Pervious Conjoint Analysis Research

In spite of the publication of more than 150 articles on conjoint analysis in top tier marketing journals (Dahan and Hauser, 2002), we know of only five published studies comparing RB and CB conjoint models. As can be seen from Table 1, all studies estimated an individual-level ratings-based model with OLS regression and an aggregate-level choice-based model. The last three studies also estimated segment-level RB and CB models using latent segment methods. The last two studies added individual-level RB and CB models estimated with HB methodology. All models were validated by predicting choices in several holdout choice sets. Some were validated at the aggregate-level by computing the mean absolute deviation (MAD) between the predicted and actual choice shares. Others were validated at the individual-level by looking at the percent of the time (hit rate) that the model correctly predicted each

individual's choices. Before comparing these validations we will look at systematic differences in the parameter estimates.

Parameter Estimates

No significant difference in the prominence effect between RB and CB models was found in Elrod et al. (1992), Oliphant et al. (1992), or Moore et al. (1998). Vriens et al. (1998) found a greater prominence effect for CB than RB models; however, Moore (2004) found a greater prominence effect for RB than CB models.

While there was no evidence for the compatibility effect in Elrod et al. (1992) or Oliphant et al. (1992), the other three studies found some support for the compatibility effect. Moore et al. (1998) found that brand name was more important in RB models. In Vriens et al. (1998), the most important quantitative (qualitative) attribute, capacity (turning filter) had a

Paper and calibration models	Ratings based conjoint validation level		Choice-based conjoint validation level	
	Individual[a]	Aggregate[b]	Individual[a]	Aggregate[b]
Elrod et al. (1992)				
OLS		.044		
Aggregate				.043
Oliphant et al. (1992)				
OLS		52.3[c]		
Aggregate				57.8
Vriens et al. (1998)				
OLS	.63			
Latent Segment	.50		.59[f]	
Aggregate	.48		.52	
Moore et al. (1998)				
OLS	.51	.049[d]		
Hierarchical Bayes	.54	.029		.028
Latent Segment	.44	.027	.62	.034
Aggregate	.43	.057	.49	.042
			.43	
Moore (2004)				
OLS	.59	.05[e]		
Hierarchical Bayes	.64	.05	.54	.07
Latent Segment	.51	.08	.42	.07
Aggregate	.46	.13	.45	.06

[a] Unless otherwise noted, the criterion is hit rate and is based on a max utility choice rule.
[b] Unless otherwise noted, the criterion is mean absolute deviation (MAD) of choice shares and is based on a max utility rule for RB models and a logit rule for CB models.
[c] Squared error between predicted and actual choices.
[d] Used a powered (rescaled) logit model to predict choice probabilities.
[e] Used an alpha model: $\alpha = 30$ for OLS, 25 for HB, 12 for latent segment, and 9 for aggregate.
[f] An accentuated number indicates a significantly higher validation than the comparable RB or CB model.

Table 1 Previous ratings-based and choice-based conjoint validation Studies

greater relative importance in CB (RB) models. Finally, Moore (2004) found that the enriched attributes, brand name and size, had significantly higher relative importances in the RB models and a comparative attribute, price, had significantly greater relative importance in the CB models.

Moore (2004) was the only study of these five about which comments on loss aversion could be made. He found loss aversion in all three attributes with more than two levels; however, in the only significant difference, the RB model had greater loss aversion than the CB model.

Validation Comparisons of RB and CB Conjoint Models

The two early studies (Elrod et al., 1992; Oliphant et al., 1992) found no significant difference in the choice share validations of individual-level RB and aggregate-level CB models.

Moore et al. (1998) found that individual-level HB CB models had significantly higher validations than comparable RB models. Similarly, in Vriens et al. (1998), segment-level CB models had higher hit rates than corresponding RB models. However, the individual- level RB model had the highest average hit rate. Additionally, Moore (2004) found that individual-level RB conjoint models generally had significantly higher validations than corresponding CB models.

Summary

The only systematic difference in parameter estimates across these studies is associated with the compatibility effect – some enriched (comparable) attributes tend to be more important in RB (CB) conjoint. The impact of level focusing is under researched.

Summarizing the validation evidence, within either CB or RB models, individual-level models have significantly higher hit rates than either segment-level or aggregate level models. When the hit rates of RB and CB models with the same level of aggregation are compared, the results are mixed, with CB models performing better slightly more often. There is no consistent difference between CB and RB models in terms of choice share validations.

There are several reasons that lead us to believe that the reevaluation of differences between choice-based and ratings-based models is necessary. The previous studies had several shortcomings. For example, in three out of five aforementioned studies that looked at this issue, the same respondents filled out both RB and CB questionnaires, creating possible order effects within person. Only two included HB models, which are the currently preferred method of performing these analyses. Finally, none was designed specifically to test for differences in level focusing effects. This study is designed to overcome all of these drawbacks.

Models

Because previous research with empirical data (see Moore (2004) for a review) finds very strong support that individual-level HB models offer better fits and validations than

comparable latent segment models, this paper focuses on HB RB and CB conjoint models (estimation details are provided in the appendix). In the RB model

$$(1)\ YR_n = X_n \hat{B}_n + e_n$$

YR_n is a (r x 1) vector containing person n's likelihood of purchase ratings, X_n is an (r x p) matrix of profile descriptions, $\hat{B}_n$ is a (p x 1) vector of person n's part worth regression weights, and e_n is a vector of error terms. The choice-based model is a logit model:

$$(2)\ P_{nil} = e^{x'nil^{\hat{B}_n}} / \sum_m e^{x'nml^{\hat{B}_n}}$$

where P_{nil} is the probability that person n chooses alternative i (out of the lth choice set whose elements are indexed by m), x_{nil} is a (p x 1) vector containing the description of the ith profile in the lth choice set seen by person n, and $\hat{B}_n$ is a (p x 1) vector of importance weights.

Individual choice predictions are based on a max utility model. Choice shares were predicted by summing individual choice probabilities over the sample. We followed Green and Krieger (1991) and estimated a decision parameter, a, that minimized the errors in choice share predictions across two three-alternative choice tasks. The following equation was used to predict individual choice probabilities with the RB models:

$$(3)\ \hat{Pr}_{ni} = \hat{YR}_{ni}^{\alpha} / \sum_{i' \in S_j} \hat{YR}_{ni'}^{\alpha} = (x'_{ni}\hat{B}_n)^{\alpha} / \sum_{i' \in S_j}(x'_{ni}\hat{B}_n)^{\alpha}$$

where $\hat{Pr}_{ni}$ is the choice probability. When α = 1, Eq. (3) is a Bradley–Terry–Luce (BTL) model. This model has been found to under predict popular choices (Green and Krieger, 1991; Moore et al., 1979; Pessemier et al., 1971). The adjustment parameter α corrects this problem.

Data

A mixed sample of upper division undergraduate business majors and MBA students from a US University was used in this study. Laptop computers were selected due to their importance to the respondent pool. This sample is comprised of knowledgeable mainstream PC users. Ninety-four percent used a PC at least once a day and 83% used one multiple times a day. Everyone used a PC for email, surfing the Web, and word processing, 93% used a spreadsheet, 66% used them for data analysis, and 46% used them for games. Ninety-six percent owned their own PC; the others accessed one at school and/or work.

Discussions with people demographically similar to the respondents, a review of computer magazines and websites, and several. Discussions with people demographically similar to the respondents, a review of computer magazines and websites, and several pre-test questionnaires generated the attributes and levels used in the study. Laptop computers were described in terms of (1) brand name: Dell, Toshiba, or WinBook; (2) Microprocessor: Intel Pentium M, Intel Celeron, or AMD Athlon; (3) Screen Size and Weight: 1200 screen and 4.5 lbs, 1400 screen and 6 lbs, or 1500 screen and 7.5 lbs; (4) Pointing device: eraserhead, touchpad, or trackball; (5) RAM (memory): 256, 512, or 768 MB; and (6) price: $999, $1199, or $1399.

Additionally, respondents were to assume that all computers had a 40 GB hard drive, a combination DVD/24_CD-R drive, wireless modem, a 3-year warranty, Windows XP Home Edition and Microsoft Office 2003 Small Business Edition.

Each level was explained. For example:

Screen size and weight. Screen size is measured diagonally in inches. Weight is typical traveling weight. There are three different combinations. They will be labeled in terms of screen size and weight, but will also differ in terms of thickness and keyboard size.

a) First is a 1200 screen and 4.5 lbs. weight – typically called an ultraportable. It has a thickness of 100. Its keyboard is 90% of the size of a desktop keyboard.
b) Next is a 1400 screen and 6 lbs. weight. It has a thickness of 1.2500. Its keyboard is 95% of the size of a desktop keyboard.
c) The final is 1500 screen and 7.5 pound weight. It is 1.500 thick. It has a keyboard that is at least 97% of the size of a desktop keyboard.

Brand name, microprocessor, and pointing device are enriched attributes and are expected to be more important in the ratings-based task. Screen size and weight, RAM, and price are easy to compare and are expected to be more important in the choice-based task.

Approximately one half of the respondents (98) received a RB questionnaire. They read a description of the attributes and levels used in the study and were given an example of a profile and the associated ratings task. Then they were presented eighteen profiles drawn from a fractional 3^6 design. They were instructed to look at all eighteen profiles to get a feel for the range of alternatives and put an ''X'' in the two or three profiles which they would be least likely to purchase and circle the two or three profiles that they would be most likely to purchase. Then they rated all profiles on a 0–10 likelihood of purchasing scale. In a second task, which was used to estimate a, they were asked to choose one of three laptops or elect to keep on shopping[5] in each of two choice tasks. Finally, they were given eight sets of four laptops and asked to indicate the one they would most likely purchase or if they would elect to keep on shopping.

The other group (n = 95) performed a CB conjoint task. They received the same description of product attributes and levels and a warm up choice task containing three alternatives as well as the option of continuing to shop. Then they received 18 choice sets, which included three laptops as well as the option to continue to shop. An initial group of choice sets was created from a cyclical design based on a 3^6 fractional factorial design, which had level balance, orthogonality, and minimal overlap. It was then modified through attribute

---

[5] Choice experiments typically employ an alternative of ''choose none'', ''stay with current product'', or, ''keep on shopping'' so the respondents are not forced to choose any of the alternatives if they are insufficiently attractive. This allows one to model changes in the size of the overall market.

relabeling to create a more utility balanced design (Huber and Zwerina, 1996).[6] They received the same set of eight calibration choice sets as the ratings-based respondents.

Results

Calibration

Table 2A compares the relative importance of the attributes in the study. An attribute's importance is defined as the difference in an individual's utility for the most and least preferred levels. These attribute importances are normalized within person.

The compatibility effect predicts that brand name, microprocessor, and pointing device would be more important in RB and that size, RAM, and price would be more important in the CB conjoint models. This was true for brand name and price. Contrary to expectations, RAM had a significantly higher importance in the RB than the CB model.

The bottom portion of Table 2A shows the relative importance of each individual's most important, second most important, and the sum of the first and second most important attributes. Both models show a strong prominence effect as the two most important attributes account for about 60% of the overall importance (compared to 33% with an equal weights model); however, there are no significant differences between models.

	RB	CB
Relative importance		
Brand	**0.167**	0.109[a]
Microprocessor	0.126	0.137
Screen size/wt	0.137	0.160
Pointing device	0.186	0.208
RAM (memory)	**0.235**	0.196[b]
Price	0.150	**0.190**[b]
Relative importance of two most important attributes		
Most important	0.381	0.379
Second most important	0.224	0.230
Sum of first two	0.604	0.609

[a] Difference significant at $p = .001$.
[b] Difference significant at $p = .05$.

Table 2A Relative attribute importance

Following Huber et al. (2002) we measured an individual's loss aversion for a given attribute by dividing the utility difference between the mid and least preferred levels by the total utility range (the attribute's importance). See Table 2B. This gives the proportion of the total utility range that is due to the difference between the two bottom levels. A proportion greater than .5 indicates diminishing returns to higher levels, or equivalently, loss aversion for

---

[6] A third group of respondents performed a CB conjoint task based on a utility balanced design that was achieved through relabeling and swapping. It produced similar results to this group, so it is not reported in order to keep the discussion focused on judgment versus choice issues. These results are available from the authors.

the lowest level. For example, the average brand preference order was Dell, Toshiba, and WinBook. Loss aversion indicates that there is a bigger gain in preference between WinBook and Toshiba than between Toshiba and Dell. Pointing device did not fit this pattern as there is a greater difference in preference between touchpad, usually the most preferred level, and the mid preference level (usually eraser head) than there is between the mid and least preferred levels (usually tracker ball). It is possible that touch pad is viewed as the reference level with regard to pointing device and anything else is viewed as creating a substantial loss.

	Brand	Processor	Screen size/ weight	Pointing device	RAM	Price
Ratings-based conjoint	0.578	0.518	0.594	0.404	0.585	0.537
Choice-based conjoint	0.512	0.499	0.553	0.442	0.665[b]	0.706[a]

[a] Difference significant at $p = .001$.
[b] Difference significant at $p = .05$.
[*] Figures of .5 or above indicate presence of loss aversion.

Table 2B Loss aversion = $(V_{mid} - V_{low})/(V_{high} - V_{low})$

While some level of loss aversion is generally observed, there is little systematic difference between RB and CB models. Two comparable attributes price ($p < .001$) and RAM ($p < .05$) exhibit greater loss aversion in CB than RB models.

In summary, Tables 2A and 2B indicate little systematic difference between ratings- and choice-based conjoint parameters. There is no evidence of a difference in the prominence effect. RAM (comparable) and pointing device (enriched) are the two most important attributes for all models. The third most important attribute is the enriched attribute, brand name, in the RB model and the comparable attribute, price, in the CB model, indicating weak compatibility effects. At least some loss aversion is exhibited for all attributes except pointing device. There is little pattern to the differences in loss aversion, but what there is seems to be related to the compatibility effect.

Individual-level Validations

Table 3A compares individual-level hit rate validations for these models across the eight holdout sets of four laptops and the option to keep on shopping. First, comparing within either RB or CB models, the individual-level parameters estimated with HB procedures produce significantly higher hit rates than either the segment-level (latent segment) or aggregate-level models, which is consistent with previous research. Focusing on the individual-level models, there are no significant differences between the HB and OLS ratings-based models. However, the individual-level CB model has significantly hit rates than either RB model.

Specifically, when predicting choices when the ''keep on shopping'' alternative is included, the HB CB model validates significantly higher than the HB RB model (.65 versus .50 hit rates). This is expected as the RB models never predicted that a person would keep on

	Including "Keep on Shopping"	Excluding "Keep on Shopping"
Ratings-based conjoint		
OLS	0.51	0.59
Hierarchical Bayes	0.50	0.58
Latent segment (7)	0.44	0.50
Aggregate	0.38	0.45
Choice-based conjoint		
Hierarchical Bayes	0.65	0.67
Latent segment (6)	0.48	0.50
Aggregate	0.38	0.44

[a] Differences of .049 are generally significant at .05 and differences of .064 or greater are generally significant at .01 level. Higher values represent stronger validations.

Table 3A Model validation on individual hit rates

shopping, which occurred about 15% of the time. Next, to create a more level playing field for the RB models and provide a better view of parameter differences, validations are carried out only on those times when a person chose one of the four laptops rather than to keep on shopping. Again, the HB CB model validates significantly higher (.67 versus .58), but the difference is less. When the keep on shopping alternative is included, the segment-level CB model validates significantly higher than the corresponding RB model (.48 versus .44), but there are no significant differences when the keep on shopping alternative is excluded (.50 versus .50) nor between any of the aggregate-level models.

Choice share validations

   Choice shares are validated in terms of MAD of predicted versus actual choice shares over all alternatives and choice sets. Choice shares are predicted using either a powered BTL (i.e., Eq. 3) with ratings-based conjoint or a logit model with choice-based conjoint. These results are presented in Table 3B.

Model	Alpha	Including "Keep on Shopping"	Excluding "Keep on Shopping"
Ratings-based			
OLS	34	.069	.033
Hierarchical Bayes	9	.076	.043
Latent segment (7)	9	.086	.054
Aggregate	9	.124	.091
Choice-based			
Hierarchical Bayes		.036	.032
Latent segment (6)		.041	.061
Aggregate		.041	.065

[a] Differences of .02 are generally significant at .05 and differences of .03 or greater are generally significant at .01 level. Lower numbers represent stronger validations.

Table 3B Choice share validations in terms of MAD

First, within either the RB or CB models, again, the individual level models validate better than segment or aggregate models. Second, when the keep on shopping choices are included, which are never predicted by the RB models, the CB models have significantly higher validations (a MAD of 0.036 versus 0.076 for the two HB models). However, when the keep on shopping choices are eliminated, there are no significant differences between the individual- or segment-level RB and CB models (e.g., a MAD of 0.032 versus 0.043 for the two HB models).

Optimal product profiles

Finally, we compared each model's share maximizing profiles. We chose the eight products from the first two choice sets as the competitive set, designated the brand name to be Dell then used the method of complete enumeration to find the share-maximizing product profiles using the parameters estimated from both models. The profile (Pentium chip, 14"/ 6lbs, touch pad, 768 RAM, and $999) that produced the highest share for the ratings-based conjoint produced the second highest share for the choice-based conjoint model.

Rather than looking at just the one share maximizing profile Table 4 shows the number of times that different attribute levels were found in the ten share maximizing profiles. While there was not complete agreement across these methods, touch pad was in all ten, Pentium in nine, 768 RAM in eight, $999 in six, and 14"/ 6lbs in at least four profiles for both models. The biggest disagreement in the top ten profiles was in the size / weight attribute (of the five attributes varied in the search, only microprocessor had a lower average importance). The RB found more top ten profiles with 14"/6lbs size, while the CB method found more with 15"/ 7.5lbs. They also differed in the second most popular size: RB - 12"/ 4.5 lbs and relabeled CB – 14"/6lbs. Overall, the methods produced similar share maximizing profiles. There were no significant differences in the forecast shares of these two models for the 20 top predicted RB sales models.

	Processor		Size/weight			Pointing device	RAM		Price		
	Pent.	AMD	14"/6 lbs	12"/4.5 lbs	15"/7.5 lbs	Touch pad	768	512	$999	$1199	$1399
Ratings-based conjoint	9	1	5	4	1	10	8	2	6	2	2
Choice-based conjoint	9	1	4	1	5	10	8	2	6	2	2

Table 4 Top ten share maximizing profiles with this option

Summary and conclusions

Operations management scholars are increasingly using both ratings-based and choice-based conjoint analysis to explore a variety of research topics. Previous empirical research comparing validations within either RB or CB models has found strong and consistent support for the use of individual-level HB models over both segment- and aggregate-level models (see Moore, 2004 and the studies reviewed therein). Therefore, this paper focuses primarily on a comparison of HB RB and HB CB models. Consistent with this previous research, our validation results did show a substantial gain for the use of individual-level models. Therefore, we strongly advocate their use.

It is well-recognized that preferences are not invariant over contexts and elicitation methods (e.g., Payne et al., 1992). In particular, there is substantial theoretical and empirical support in the behavioral decision theory (BDT) literature for the prominence, compatibility and level-focusing effects. For example, Fischer et al. (1999) argue that the prominence effect, the tendency to give more weight to important attributes in some preference elicitation tasks than others, is one of the strongest preference biases in the decision making literature. Compatibility effects, or the tendency to give more weight to attributes that require smaller transformations in a preference task, were found to be substantial by Hsee (1996) and Nowlis and Simonson (1997) where enriched (comparable) attributes were more important in judgment (choice) tasks. Finally, level-focusing is the tendency to focus on certain attribute levels more than others (e.g., Kahneman and Tversky, 1979).

This study, as well as those reviewed, does not indicate a strong systematic difference between the weights estimated with these two methods that could be attributed to earlier suggested prominence, compatibility, or level focusing effects. The general consensus of these studies is that (1) while both models find some attributes to be much more important than others the difference in the prominence effect between RB and CB models is typically small, (2) RB (CB) conjoint analysis are likely to give slightly greater importance to a subset of the enriched (comparable) attributes like brand name (price) and (3) based on this study, one might expect marginally greater level focusing on some comparable attributes with CB conjoint analysis. However, the impact of these differences is relatively small: in this study, they had only a minor impact on the optimal profiles. Across all studies there were only small differences in choice share validations when the ''keep on shopping'' choices were excluded. This indicates that one would generally expect the parameters estimated from RB and CB conjoint analyses to be similar,4 but not identical (see also Conlon et al., 2000). This means that the findings with respect to the most important attributes and most preferred levels within attribute will usually be quite similar, leading to similar optimal designs.

The lack of systematic differences attributed to the above mentioned effects in conjoint studies could be due to the larger number of product profiles and attributes involved in the conjoint tasks. Conjoint studies use more complex stimuli – typically six to ten attributes in contrast to just two attributes used in many BDT studies. In addition, we use warm up tasks, and conjoint respondents make repeated choices from the same product class where as the typical behavioral decision theory studies have only one choice from a given product class. Lack of strong differences in prominence, compatibility, and level-focusing effects (regardless of the reason explaining their absence), allows us to conclude that these effects will not generally produce large systematic differences between the part-worths estimated in CB and RB studies.

The biggest systematic difference may not be in the parameter estimates, but in CB conjoint's greater ability to model whether a customer makes a purchase from a product class or not in terms of either individual hit rates or choice share validations (i.e., see the 'Including ''Keep on Shopping'' columns in Tables 3A and 3B). Inclusion of the ''Keep on shopping'' option

allows one to model market expansion or contraction based on the attractiveness of the competitive alternatives. While reservation price models have been developed for RB conjoint (e.g., Jedidi and Zhang, 2002), we do not know of any rigorous comparisons with CB approaches. This suggests CB models may be more appropriate when it is important to model changes in market size by giving respondents the option to not purchase anything. Additionally, this will lead to slightly lower profit maximizing prices with CB conjoint models.

A currently active area in conjoint analysis involves trying to predict either competitor or supply chain reactions. These methods usually focus on finding a Nash equilibrium – a point at which none of the competitors has an incentive to move, e.g., a set a competitive prices where none of the competitors would be better off by changing his/her price. One approach to solving this problem is to simulate a series of optimal decisions by each of the competitors until none make further changes (e.g., Green et al., 2005). The more typical approach involves finding the equilibrium by differentiating all the competitive profit equations by the variable of interest, e.g., price, to find the equilibrium prices. CB conjoint models seem to be particularly well suited to this second method of analysis (e.g., Luo et al., 2007).

Both RB and CB methods have a successful history of application and the continuing popularity of RB conjoint, as evidenced by books such as Krieger et al. (2004), suggests both can be appropriate choices. However, our findings that HB CB conjoint models had higher individual-level validations than comparable RB models, causes us to recommend CB models when individual-level predictions are desired. Additionally, the higher accuracy of the individual level predictions gives us more confidence in the choice-based models. Additional reasons for choosing CB conjoint analysis include the ability to model changes in market size and its impact on equilibrium prices. Finally, the availability of software (e.g., Sawtooth Software, 2000–2005) has made it much easier to design and conduct CB conjoint studies.

This study, along with the other articles in this research stream, strongly suggests that in traditional conjoint tasks, the parameter estimates produced by RB and CB conjoint models are likely to be quite similar. While additional direct comparison studies may be less valuable, we see a number of exciting future research opportunities relevant to these two model types and their role within emerging conjoint analysis frameworks. For example, one area involves ways to modify traditional RB models for situations where consumers encounter a large number of product profiles (cf., Gilbride and Allenby, 2004 and Yee et al., 2007). Ratings-based tasks encourage consumers to engage in compensatory decision making, which may not be well equipped for capturing non-compensatory decisions and elimination strategies used by consumers in complex decision situations. A second important area for both RB and CB is how to cope with larger numbers of attributes and levels within attribute.

Another area involves the use of incentives-aligned (IA) conjoint tasks, where consumers are rewarded for participating in the experiment based on either their (a) choices or (b) contingent valuation responses. Ding et al. (2005) find that choice-based IA conjoint results in better out-of-sample validations than contingent valuation responses (which are similar to the

traditional RB responses in that both require the respondents to assign a numerical value to the alternative; however, in contingent valuation, the respondent is asked to assign a dollar value rather than a preference rating). These findings suggest that various types of preference elicitation techniques may be more or less successful when used with new conjoint analysis approaches. More research is needed to identify and explore these potential differences.

Appendix A. Estimation details

Ratings-based models are typically estimated with regression:

$$(1)\ YR_n = X_n \hat{B}_n + e_n$$

where YR_n is a mean-centered (r x 1) vector containing person n's likelihood of purchase ratings, X_n is an (r x p) matrix of profile descriptions, $\hat{B}_n$ is a (p x 1) vector of person n's part worth regression weights, and e_n is a vector of error terms. We estimated an individual OLS model by running a separate regression for each person, an individual-level HB model (details below), a segment-level model (details below), and an aggregate model where preferences were pooled over all individuals.

Choice-based models are typically estimated with logit models.

$$(2)\ P_{nil} = e^{x'nil^{\hat{B}_n}} / \sum_m e^{x'nml^{\hat{B}_n}}$$

where P_{nil} is the probability that person n chooses alternative i (out of the lth choice set whose elements are indexed by m), x_{nil} is a (p x 1) vector containing the description of the ith profile in the lth choice set seen by person n, and $\hat{B}_n$ is a (p x 1) vector of importance weights.

A1. Hierarchical Baynesian models

Respondent part-worth heterogeneity is assumed to be distributed multivariate normal. $\beta_n \sim N\ (\theta, \Lambda)$, where θ is a (p x 1) vector of mean part worths, and Λ is a (p x p) positive definite matrix that follows an inverse Wishart$_p$ (η_n, Δ_n) distribution with prior parameters $\eta_0 = p$ and$\Delta_0 = pI$, and $\theta \sim N\ (\theta_n, V_n)$ with prior parameters $\theta_0 = 0$ and$V_0 = 100I$. Additionally, in the RB model, $\sigma_i^2 \sim$ Inverse gamma $(\alpha_n/2, \psi_n/2)$ with a prior distribution IG $(\alpha_0/2, \psi_0/2)$, $\alpha_0 = \psi_0 = 2$. In both cases, 10,000 draws were made, the first 5000 as a ''burn in'' and the last 5000 for the use in parameter estimation.

A.2. Latent segment models

Latent segment models were estimated with an expectation–maximization (EM) algorithm. For a given number of segments, respondents were randomly placed in segments, models were estimated for each segment and respondents were reassigned to the segment for which they had the highest probability of membership, segment models were re-estimated and people reassigned until no further changes took place. This procedure is repeated with different numbers of segments and the appropriate number of segments is chosen with the CAIC, Consistent Akaike Information Criterion, which penalizes the better fits more complex models

(models with more segments in this case) for their greater number of parameters. After an appropriate segment-level model is chosen, respondent weights are a weighted average of the segment weights and the probability of segment membership as the weight.

References

Allenby, G.M., Lenk, P.J., 1994. Modeling household purchase behavior with logistic normal regression. Journal of the American Statistical Association 89, 1218–1231.

Bendoly, E., Donohue, K., Schultz, K., 2006. Behavior in operations management: Assessing recent findings and revisiting old assumptions. Journal of Operations Management 24 (6), 737–752.

Bettman, J.R., Luce, M.F., Payne, J.W., 1998. Constructive consumer choice processes. Journal of Consumer Research 25 (3), 187–217.

Butler, J.C., Dyer, J.S., Jia, J., Tomak, K., 2008. Enabling e-transactions with multiattribute preference models. European Journal of Operational Research 186 (2), 748–765.

Conlon, B., Dellaert, B.G.C., van Soest, A., 2000. Combining and comparing consumers' stated preference ratings and choice responses. Research Paper 2000-119. Center for Economic Research 2000. http://publishing.eur.nl/ir/repub/asset/11114/CombiningAndComparing.pdf (last accessed 5/21/2008).

Dahan, E., Hauser, J.R., 2002. Managing a dispersed product development process. In: Weitz, B., Wensley, R. (Eds.), The Handbook of Marketing. Sage, Beverley Hills, CA, pp. 179–222.

Day, J.M., Venkataramanan, M.A., 2006. Profitability in product line pricing and composition with manufacturing commonalities. European Journal of Operational Research 175 (3), 1782–1797.

Ding, M., Grewal, R., Liechty, J., 2005. Incentives-aligned conjoint analysis. Journal of Marketing Research 42 (1), 67–82.

Ding, X., Verma, R., Iqbal, Z., 2007. Self-service technology and online financial service choice. International Journal of Service Industry Management 18 (3), 246–268.

Elrod, T., Louviere, J., Davey, K., 1992. An empirical comparison of ratings-based and choice-based conjoint models. Journal of Marketing Research 24 (3), 368–377.

Fischer, G.W., Carmon, Z., Ariely, D., Zauberman, G., 1999. Goal-based construction of preferences: Task goals and the prominence effect. Management Science 45 (8), 1057–1075.

Fischer, G.W., Hawkins, S.A., 1993. Strategy compatibility, scale compatibility, and the prominence effect. Journal of Experimental Psychology: Human Perception and Performance 19 (3), 580–597.

Gilbride, T., Allenby, G., 2004. A choice model with conjunctive, disjunctive, and compensatory screening rules. Marketing Science 23 (30), 391–406.

Goldstein, W.M., Einhorn, H.J., 1987. Expression theory and the preference reversal phenomena. Psychological Review 94 (2), 236–254.

Goodale, J.C., Verma, R., Pullman, M.E., 2003. A market utility-based model for capacity scheduling in mass services. Production and Operations Management 12 (2), 165–185.

Green, P.E., Carroll, J.D., DeSarbo, W.S., 1981. Estimating choice probabilities in multiattribute decision making. Journal of Consumer Research 8 (1), 76–84.

Green, P.E., Krieger, A.M., 1991. Segmenting markets with conjoint analysis. Journal of Marketing 55 (4), 20–31.

Green, P.E., Krieger, A.M., 1992. Modeling competitive pricing and market share: Anatomy of a decision support system. European Journal of Operational Research 60 (1), 31–44.

Green, P.E., Krieger, A.M., Wind, Y., 2005. Buyer choice simulators, optimizers, and dynamic models. In: Wind, Y., Green, P., (Eds.), Market Research and Modeling: Progress and Prospects: A Tribute to Paul E. Green. Springer Science + Business Media Inc., pp. 169–199.

Hsee, C.K., 1996. The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. Organizational Behavior and Human Decision Processes 67 (3), 247–257.

Huber J., 1987. Conjoint analysis: How we got there and where we are. In: 1987 Sawtooth Software Proceedings. Ketchum, ID: Sawtooth Software, pp. 237–253.

Huber, J., Ariely, D., Fischer, G., 2002. Expressing preferences in a principal-agent task: A comparison of choice, rating, and matching. Organizational Behavior and Human Decision Processes 87 (1), 66–90.

Huber, J., Zwerina, K., 1996. The importance of utility balance in efficient choice set designs. Journal of Marketing Research 33 (3), 307–317.

Jedidi, K., Zhang, Z.J., 2002. Augmenting conjoint analysis to estimate consumer reservation price. Management Science 48 (10), 1350–1368.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision making under risk. Econometrica 47 (2), 18–36.

Karmarkar, U.S., 1996. Integrative research in marketing and operations management. Journal of Marketing Research 33 (2), 125–133.

Kim, S.-H., Srinivasan, V., 2006. A multiattribute model of the timing of buyers' upgrading to improved versions of high technology products. Research Paper No. 1720(R), September, Graduate School of Business, Stanford University.

Krieger, A.M., Green, P.E., Wind, Y.J., 2004. Adventures in conjoint analysis: A practitioner's guide to trade-off modeling and applications, 2004 Draft Manuscript. http://marketing.wharton.upenn.edu/people/faculty/green/green_monograph.cfm (accessed 2/21/2008).

Krishnan, V., Ulrich, K.T., 2001. Product development decisions: A review of the literature. Management Science 47 (1), 1–21.

Li, S., Madhok, A., Plaschka, G., Verma, R., 2006. Supplier-switching inertia and competitive asymmetry: A demand side perspective. Decision Sciences 37 (4), 547–576.

Lenk, P.J., DeSarbo, W.S., Green, P.E., Young, M.R., 1996. Hierarchical Bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental designs. Marketing Science 15 (2), 173–191.

Luo, L., Kannan, P.K., Basharati, B., Azarm, S., 2005. Design of robust new products under variability: Marketing meets design. Journal of Product Innovation Management 22 (2), 177–192.

Luo, L., Kannan, P.K., Ratchford, B.T., 2007. New product development under channel acceptance. Marketing Science 26 (2), 149–163.

Mankila, M., 2004. Retaining students in retail banking through price bundling: Evidence from the Swedish market. European Journal of Operational Research 155 (2), 299–316.

Michalek, J., Feinberg, F.M., Papalambros, P.Y., 2005. Linking marketing and engineering product design decisions via analytic target cascading. Journal of Product Innovation Management 22 (1), 42–62.

Moore, W.L., 2004. A cross-validity comparison of ratings-based and choice-based conjoint analysis models. International Journal of Research in Marketing 21 (3), 299–312.

Moore, W.L., Gray-Lee, J., Louviere, J.J., 1998. A cross-validity comparison of conjoint analysis and choice models at different levels of aggregation. Marketing Letters 9 (2), 195–208.

Moore, W.L., Louviere, J.J., Verma, R., 1999. Using conjoint analysis to help design product platforms. Journal of Product Innovation Management 16 (1), 27–39.

Moore, W.L., Pessemier, E.A., Little, T.E., 1979. Predicting brand choice behavior: A marketing application of the Schönemann and Wang unfolding model. Journal of Marketing Research 16 (2), 203–210.

Morgan, L.O., Daniels, R.L., Kouvelis, P., 2001. Marketing/manufacturing tradeoffs in product line management. IIE Transactions 33 (11), 949–962.

Nair, S.K., Tarasewich, P., 2003. A model and solution method for multi-period sales promotion design. European Journal of Operational Research 150 (3), 672–687.

Natter, M., Feurstein, M., 2002. Real world performance of choice-based conjoint models. European Journal of Operational Research 137 (2), 448–458.

Nowlis, S.M., Simonson, I., 1997. Attribute task comparability as a determinant of consumer preference reversals. Journal of Marketing Research 34 (2), 205–218.

Oliphant, K., Eagle, T.G., Louviere, J.J., Anderson, D., 1992. Cross-task comparison of ratings-based and choice-based conjoint. In: Metegrano, M. (Ed.), 1992 Sawtooth Software Conference Proceedings, pp. 383–404.

Orme, B.K., Alpert, M.I., Christensen, E., 1997. Assessing the validity of conjoint analysis – continued. Sawtooth Software Research Paper Series 1997. http://www.sawtoothsoftware.com/download/techpap/assess2.pdf (accessed 2/21/2008).

Payne, J.W., Bettman, J.R., Johnson, E.J., 1992. Behavioral decision research: A constructive processing perspective. Annual Review of Psychology 43, 87–131.

Pessemier, E.A., Burger, P., Teach, R., Tigert, D., 1971. Using laboratory brand preference scales to predict consumer brand purchases. Management Science 17 (6), 371–385.

Pullman, M.E., Moore, W.L., 1999. Optimal service design: Integrating marketing and operations perspectives. International Journal of Service Industry Management 10 (2), 239–260.

Sawtooth Software (2000–2005). The CBC/HB system for hierarchical Bayes estimation, Version 4.0 technical paper, Sawtooth Software Technical Paper Series, Sequim, WA. http://www.sawtoothsoftware.com/download/techpap/hbtech.pdf (accessed 2/21/2008).

Schmidt, G.M., Porteus, E.L., 2000. The impact of an integrated marketing and manufacturing innovation. Manufacturing and Service Operations Management 2 (4), 317–336.

Scholl, A., Manthey, L., Helm, R., Steiner, M., 2005. Solving multiattribute design problems with analytic hierarchy process and conjoint analysis: An empirical comparison. European Journal of Operational Research 164 (3), 760–777.

Tarasewich, P., McMullen, P.R., 2001. A pruning heuristic for use with multisource product design. European Journal of Operational Research 128 (1), 58–73.

Tatikonda, M.V., Montoya-Weiss, M.M., 2001. Integrating operations and marketing perspectives of product innovation: The influence of organizational process factors and capabilities on development performance. Management Science 47 (1), 151–172.

Turksen, I.B., Willson, I.A., 1995. A fuzzy set model for market share and preference prediction. European Journal of Operational Research 82 (1), 39–52.

Tversky, A., Sattath, S., Slovic, P., 1988. Contingent weighting in judgment and choice. Psychological Review 95 (3), 371–384.

Ülengin, B., Ülengin, F., Güvenç, Ü, 2001. A multidimensional approach to urban quality of life: The case of Istanbul. European Journal of Operational Research 130 (2), 361–374.

Verma, R., Louviere, J.J., Burke, P., 2006. Using market-utility-based approach to designing government and public services: Case illustration from United States Forest Service. Journal of Operations Management 24 (4), 407–416.

Verma, R., Pullman, M., 1998. An analysis of the supplier selection process. Omega: The International Journal of Management Science 26 (6), 739–750.

Verma, R., Thompson, G.M., Moore, W.L., Louviere, J.J., 2001. Effective design of products/services: An approach based on integration of marketing and operations management decisions. Decision Sciences 32 (1), 165–194.

Verma, R., Plaschka, G.R., Hanlon, B., Livingston, A., Kalcher, K., 2008. Predicting customer choice in services using discrete choice analysis. IBM Systems Journal 47 (1), 179–191.

Victorino, L., Verma, R., Plaschka, G., Dev, C., 2005. Service innovation and customer choices in hospitality industry. Managing Service Quality 15 (6), 555–576.

Vriens, M., Oppewal, H., Wedel, M., 1998. Rating-based versus choice-based latent class conjoint models: An empirical comparison. Journal of the Market Research Society 40 (3), 237–248.

Weitz, B., Wright, P., 1979. Retrospective self-insight on factors considered in product evaluations. Journal of Consumer Research 6 (3), 280–294.

Yano, C., Dobson, G., 1998. Profit-optimizing product line design, selection and pricing with manufacturing cost consideration. In: Ho, T.-H., Tang, C.S. (Eds.), Product variety management: Research advances. Kluwer Academic Publishers, Boston, pp. 145–173.

Yee, M., Dahan, E., Hauser, J.R., Orlin, J., 2007. Greedoid-based noncompensatory inference. Marketing Science 26 (4), 532–549.