Application and Test of Web-based Adaptive Polyhedral Conjoint Analysis

by

Ely Dahan

John R. Hauser

Duncan Simester

and

Olivier Toubia

October 13, 2001

Ely Dahan is an Assistant Professor of Marketing and Management Science, E56-323, Sloan School of Management, Massachusetts Institute of Technology, 38 Memorial Drive, Cambridge, MA 02142-1307, (617) 253-0492, edahan@mit.edu.

John Hauser is the Kirin Professor of Marketing and Management Science, E56-314, Sloan School of Management, Massachusetts Institute of Technology, 38 Memorial Drive, Cambridge, MA 02142-1307, (617) 253-2929, jhauser@mit.edu.

Duncan I. Simester is an Associate Professor, Sloan School of Management, Massachusetts Institute of Technology, E56-305, 38 Memorial Drive, Cambridge, MA 02142, (617) 258-0679, fax (617) 253-7597, simester@mit.edu.

Olivier Toubia is a graduate student at the Marketing Group and the Operations Research Center, Massachusetts Institute of Technology, E56-345, 38 Memorial Drive, Cambridge, MA 02142, (617) 253-0159, fax (617) 253-7597, toubia@mit.edu.

Application and Test of Web-based Adaptive Polyhedral Conjoint Analysis

Acknowledgements

This research was supported by the Sloan School of Management, the Center for Innovation in Product Development at MIT , and the eBusiness Center at MIT. This paper may be downloaded from http://mitsloan.mit.edu/vc. That website also contains (1) open source code to implement the methods described in this paper, (2) demonstrations of web-based questionnaires used to collect data for the tests described in this paper, and (3) related papers and demonstrations of web-based interviewing methods. All authors contributed fully and synergistically to this paper.

We wish to thank Limor Weisberg for creating the graphics used on the web-based questionnaire and Rob Hardy, Adler Kim, and Bryant Lee who assisted in the development of the web-based questionnaire. We gratefully acknowledge the contribution of Robert M. Freund who proposed the use of the analytic center and approximating ellipsoids for the math programs required for question selection and partworth estimation. The questionnaire benefited from comments by numerous pretesters at the MIT Operations Research Center, Analysis Group Economics, Applied Marketing Science, the Center for Innovation in Product Development, and Dražen Prelec's market measurement class. Our thanks to Brennan Mulligan of Timbuk2 Designs, Inc. for supplying the laptop computer bags used in our experiments.

Application and Test of Web-based Adaptive Polyhedral Conjoint Analysis

Abstract

In response to the need for more rapid and iterative feedback on customer preferences in the new product development process, researchers are developing new web-based conjoint analysis methods that promise to reveal preference information for more product features with fewer questions. In this paper we apply and test one such method, based on "interior-point" math programming methods, that adapts questions for each respondent and provides preference estimates for each respondent. The basic "polyhedral" method is benchmarked with an adaptive standard, adaptive conjoint analysis (ACA), and a non-adaptive standard, an efficient fixed design. To standardize comparisons among methods, the core conjoint task in all three methods is graded-paired (metric) paired-comparison questions.

Over 300 respondents were randomly assigned to different experimental conditions, each representing a method, and completed a web-based conjoint exercise. After a filler task respondents were given $100 that they used to make a purchase from a pareto choice set of five new-to-the-market laptop computer bags. The respondents received their chosen bag together with the difference in cash between the price of their chosen bag and the $100.

We compare the methods on both internal and external validity. Internal validity is evaluated by comparing how well the methods predict several holdout conjoint questions. External validity is evaluated by comparing how well the different conjoint methods predict the respondents' selection from the choice set of five bags.

The results reveal a remarkable level of consistency in the relative performance of the different methods across the different validation tasks. The polyhedral method was consistently more accurate than both the ACA and fixed methods. However, even better performance was achieved by combining (post hoc) different features of each method to create a range of hybrid methods. Additional analyses evaluate the robustness of the predictions with respect to question order and explore alternative estimation methods such as Hierarchical Bayes and balanced estimation. At the time of the test, the bags were prototypes. Based, in part, on the results of this study these bags are now available commercially.

Recent Developments

Preference measurement has a long history in marketing. Firms routinely rely upon various forms of conjoint analysis, self-stated importances, and voice-of-the-customer methods to design new products, expand product lines, and set segmentation strategies (e.g., Cattin and Wittink 1982; Green and Srinivasan 1990; Wittink and Cattin 1989). Until recently, preference measurement could safely be called a mature technology with a rich history of applications, improvements, simulation tests, and empirical tests. However, in the past few years three developments have caused researchers to refocus on this fertile area of research.

The first development is a change in the product development (PD) process. Business-to-business and complex durable goods manufacturers are adopting consumer-packaged-good methods. Because these products are typically complex with many features, PD teams are demanding preference measurement methods than can handle more customer needs and provide prioritizations among substantially more features (Eppinger 1998; Eppinger, et. al. 1994; Ulrich and Eppinger 2000). At the same time concurrent engineering, rapid web-based product design, and "spiral" processes view preference measurement as iterative (Cusumano and Selby 1995; McGrath 1996; Smith and Reinertsen 1998; Tessler, Wada and Klein 1993). PD teams routinely screen large numbers of features, then focus, remeasure, and iterate. Accuracy need only be sufficient to separate the winning features from the losing features, shifting the tradeoff toward lower research costs per iteration. PD teams are more willing to sacrifice some precision in the measurement of preferences in order to handle more features.

The second development is the advent of the Internet as a means to collect data. Web-based panels with large numbers of respondents provide the potential for very rapid and reliable feedback (Buckman 2000; Gonier 1999; Nadilo 1999; Willkie, Adams, and Girnius 1999). A conjoint analysis study, once designed, can be fielded nationally or internationally in days. Furthermore, with broadband capabilities and today's software tools, virtual concepts can be created with vivid representations, animation, context-sensitive help, and sound (Dahan and Hauser 2001; Dahan and Srinivasan 2000). In addition, the ability for computation between questions has renewed interest in new forms of adaptive questions – questions selected and created for each customer based on his or her prior answers. Adaptive conjoint analysis (ACA), long the most-widely used adaptive method, is now available as a web-based tool (www.sawtoothsoftware.com). The Internet has also brought with it a new set of limitations. In particular, web-based interviewing techniques must address reduced respondent tolerance for answering multiple questions. In central location interviewing, respondents are remarkably tolerant of multiple questions. However, the literature and commercial experience indicate much lower tolerance levels in a web context (e.g., De Angelis 2001). In our experience, interest and reliability drop off rapidly after eight to twelve repetitive questions. This creates greater demand for conjoint methods that gather more information with fewer questions.

The third development is the introduction of projective interior-point methods in math programming. Conjoint analysis has used math programming successfully for almost thirty years (Srinivasan and Shocker 1973a, 1973b), but its application to adaptive question selection is new. For large problems, optimization for individual-respondent question selection is a complex, difficult-to-solve problem, which, by traditional methods, might lead to excessive delays between questions – a situation that is unacceptable to web-based respondents. The new algorithms search the interior of a parameter space and can find near-optimal solutions extremely fast (Freund 1993; Kamarkar 1984; Nesterov and Nemiroskii 1994; Sonnevend 1985a, 1985b; Vaidja 1989). This capability is leading to new conjoint analysis methods that can provide information about many features from few questions. This represents a change in philosophy from a primarily statistical focus, to a focus on feasible heuristics. These methods do well in simulation, but have not yet been tested on real customers.

In this paper we evaluate the interior point method proposed by Toubia, Simester and Hauser (2001). For benchmarks we compare their so-called FastPace method with ACA and a fixed (non-adaptive) method (Green, Krieger and Agarwal 1991; Johnson 1991; Kuhfeld, Tobias and Garratt 1994; Orme 1999). We use a real product – a laptop computer bag worth approximately $100. After completing the various web-based conjoint questionnaires (assigned randomly), respondents in the study were given $100 to make an actual purchase from a simulated store. Respondents received their chosen bag together with the difference in cash between the price of their chosen bag and the $100.

This data enables us to compare the methods on both internal and external validity. Internal validity is evaluated by comparing how well the methods predict several holdout conjoint questions. External validity is evaluated by comparing how well the different conjoint methods predict the respondents' selection of a laptop bag that they could keep (with real cash at risk). We then explore whether there is a question-order effect, whether the performance of the "purebred" methods can be improved with hybrids that combine features from the different methods, and whether alternative estimation methods such as Hierarchical Bayes methods improve internal and external validity.

We begin with a brief review of the methods we test and then describe the research design. The core of the paper reports on and interprets the internal and external validity tests. We close with a summary and discussion of the findings. (To avoid redundancy throughout the paper we rely on many two-letter and three-letter acronyms. To help the reader, we repeat key definitions when we begin new sections.)

Three Conjoint Analysis Methods

There are many forms of data collection for conjoint analysis, but most fall into one of three categories – metric paired comparisons, choice from a set of profiles, and ranking or rating of a factorial design of profiles. Polyhedral methods can be devised for all three tasks. In this paper, we focus on the task that is most natural and common in web-based interviewing – metric paired comparison data. This is also consistent with the state-of-the-art in polyhedral methods. Because our focus is on adaptive methods using metric paired-comparison questions, we focus on two adaptive methods (ACA and FastPace) as well as a standard non-adaptive method (efficient fixed designs). We also explore hybrids and improvements based on these data collection models. As all three methods are described elsewhere, our review is brief (readers seeking more detail should consult the references).

Conjoint analysis represents a product as a profile of its features and price. Discrete levels are generally used to represent each feature, such as 50, 70 and 90 cubic feet for the cargo capacity of a new cross-over vehicle. The preference for a product is then a function of the "partworths" assigned to the levels of each feature. In this study we focus on 10 separable features of a laptop computer bag, where each feature has two levels. In our application, the chosen features satisfied preferential independence allowing us to model preference for each bag as a linear additive combination of these features (Keeney and Raiffa 1976, p. 101-105; Krantz, et. al. 1971). We recognize that in other applications features may have more than two levels and may not be separable, discrete, or additive.

Conjoint analysis describes a class of techniques that ask respondents their preferences for different product profiles and then estimate partworths based on the answers to these questions. We hold the question format constant across methods. In particular, for this empirical test, we use interval-scaled paired-comparison questions (PC) in which respondents choose among two product profiles and indicate their strength of preference. This is a common format, well suited to the web, standard in the ACA method, and shown to provide interval-scaled preferences (Hauser and Shugan (1980); Johnson 1991; Mahajan and Wind 1992; Wittink and Cattin 1989). All methods also ask initial questions to determine which level of each feature is "high" and which is "low."

The "purebred" methods vary in at least two respects: (1) the heuristics that they use to select the product profiles that respondents compare, and (2) the manner in which partworths associated with each feature are estimated. In the discussion that follows, we review the design of the product profiles and the basic estimation procedures for the three methods. We consider alternative estimation methods, such a hybrid methods, balanced methods, and hierarchical Bayes methods in later sections.

Fast Polyhedral Adaptive Conjoint Estimation (FastPace)

Fast polyhedral adaptive conjoint estimation (FP) uses recent advances in interior point methods to design questions (product profiles) and estimate partworths. Toubia, Simester and Hauser (2001) propose one implementation and test it using simulated respondents. Their method interprets the problem of estimating partworths for p different product features as searching an p-dimensional space.[1] A point in this p-dimensional space represents one set of relative importances (partworths) for the features. The goal is to identify the point in the space that best represents the true partworths for an individual respondent.

When respondents answer a metric question comparing one or more product profiles, their answers will generally be consistent with only a subset of the points in the p-dimensional space. Other points imply partworths that are not consistent with the respondent's answer. A respondent's answer is more informative if it more quickly reduces the range of points that are consistent with the answer, thus FP questions quickly reduce the set of consistent (feasible) points. The FP method approximates the current feasible (polyhedral) region using ellipsoids. By selecting question vectors that are perpendicular to the longest axis of the ellipsoid, the feasible region shrinks rapidly (and inconsistent responses are less likely). We refer interested readers to their paper, where the authors describe the details and rationale for this approach, includ-

[1] FP allows the initial space to also be constrained by prior information or other constraints. For features with more than two levels or for features that interact, FP simply expands the dimensionality of the space accordingly.

ing means to model respondent errors. (Their paper and open-source code are available at the website listed in the acknowledgments section of this paper.)

The managerial problem imposed by the recent changes in product development requires estimates of partworths with a relatively few questions – often fewer questions than there are features. This means that the researcher must often estimate the respondent's revealed partworths even though there remains a sizable region of feasible partworths that are consistent with the respondent's answers to the chosen questions. Based on research that justifies centrality as a criterion (Dawes and Corrigan 1974; Einhorn 1971; Huber 1975; Moore and Semenik 1988; Srinivasan and Park 1997), FP uses an estimate that characterizes the set of remaining feasible points by the center of gravity of the set. Because it is computationally difficult to identify the center of gravity of a polyhedron, they use the analytic center, which an excellent approximation. The analytical center can be estimated with effectively no computational delay between questions.

Tests conducted with simulated respondents suggest that the algorithm offers improved accuracy over ACA and standard fixed efficient-design methods when there are many features relative to the number of questions and when respondent wearout is an issue. Toubia, Simester and Hauser (2001) also consider a range of hybrid methods and show that they have the potential to improve performance.

Adaptive Conjoint Analysis (ACA)

Sawtooth Software's Adaptive Conjoint Analysis (ACA) has become the industry standard for adaptive questions to collect data for conjoint analysis (Green, Krieger and Agarwal 1991, p. 215). See among others: Carroll and Green 1995; Choi and DeSarbo 1994; Green and Krieger 1995; Green, Krieger and Agarwal 1991; Huber, Wittink, Fiedler and Miller 1993; Johnson 1987; Klein 1988; Orme 1999; Wittink and Cattin 1989. Although there is some variation in conclusions, academic and commercial experience suggest that ACA is often reliable, accurate, and managerially valuable.

The ACA interview consists of four sequential tasks (Klein 1988; Sawtooth 1996). Respondents (1) eliminate unacceptable levels, (2) state the relative importances of improving a product from one feature level to another (the "self-explicated" questions), (3) complete the metric paired-comparison tasks, and (4) evaluate a small set of full product profiles on a purchase intention scale. The "adaptive" task is the third task; the first task is often skipped. The ACA algorithm uses information from a respondent's earlier responses to provide current estimates of a respondent's preferences and then designs product profiles that are most nearly equal in preference, subject to constraints that ensure the overall design is nearly balanced and orthogonal. Both the intermediate estimates and the final estimates are based on a modified regression that combines the information from the self-explicated (SE) and the paired-comparison (PC) data. The intermediate regression has $p+q$ data points – one for each of p SE questions and one for each of q PC questions. Because the regression always has at least p observations, ACA can provide intermediate estimates even after a single PC question. Once the data collection is complete, a logit regression based on the intention scales (the fourth task) determines how heavily to weight the SE and PC data. For more details see Green, Krieger and Agarwal 1991 and Sawtooth (1996).

Fixed Efficient Design

To test whether the adaptive nature of the questions improves or degrades accuracy we chose as a benchmark a method in which set of PC questions is chosen in advance. Every respondent sees the same set of questions, although the order of the questions is randomized. For the fixed design we used an algorithm based on the geometric means of the eigenvalues of the information matrix as a means to maximize a D-efficiency criterion – a criterion that yields profile designs that are as balanced and orthogonal as feasible (Kuhfeld 1999; Kuhfeld, Tobias and Garratt 1994). Because the PC data are metric, the partworths are estimated with ordinary least-squares regression based on the q questions. Regression requires more data points than estimated parameters, and so the fixed method provides estimates only after q exceeds p by sufficient degrees of freedom. (In later sections we explore Hierarchical Bayes estimates that do not require q to exceed p.)

All three methods that we evaluate use metric PC questions, but only ACA requires the SE (and PI) questions. Therefore, when responses to the SE questions are more accurate, we expect an improvement in the relative performance of ACA. Consistent with this, Toubia, Simester and Hauser's (2001) simulations suggest that if the SE responses are sufficiently accurate, then ACA outperforms both the fixed and FP methods. ACA does not perform as well as the other two methods when the SE responses are noisy.

The Across-Subjects Research Design to Compare Methods

The product used in the study was an innovative new laptop computer bag that includes a removable padded sleeve to hold and protect a laptop computer. At the time of our study, this product was not yet on the market so our respondents had no prior experience with the product. Subsequent to our research and based, in part, on data from our research, the product is now available commercially from Timbuk2 (http://www.timbuk2.com). The bag includes a range of separable product features, such as the inclusion of a mobile-phone holder, side pockets, or a logo. We focused on nine product features, each with two levels, and included price as a tenth variate. In the conjoint exercise, price is restricted to two levels ($70 and $100) – the extreme price levels for the bags in both the internal and external validity tests. For all three conjoint methods, we estimated the partworths associated with prices between $70 and $100 by linearly interpolating between the partworths estimated for the $70 and $100 levels. Figure 1 offers a photograph illustrating examples of the bags. A more detailed description of the product features can be found on the website listed in the acknowledgements section of this paper.

Our focus on product features that are more or less separable supports our ability to represent preferences for profiles as an additive function of the partworths. The separable nature of the product features may also influence the accuracy of the self-explicated (SE) questions used by ACA. In particular, we expect SE responses to be more accurate in categories where customers make purchasing decisions about features separately, perhaps by choosing from a menu of features. In contrast, we expect SE responses to be less accurate for products where the features are typically bundled together, so that customers have little experience in evaluating the importance of the individual features. Thus, our choice of separable features is likely to favor ACA relative to the FP (and fixed) methods and, as such, provides a conservative test of the FP method (relative to ACA).

Figure 1. Examples of Product Category (Laptop Computer Bags)

Research Design

We begin with an overview of the research design and then discuss each element of the design in greater detail. The overview of the research design is summarized in Table 1 and the detailed research design is summarized in Table 2. Subjects were randomly assigned to one of the three conjoint methods. After completing their respective conjoint tasks, the respondents were all presented with the same validation exercises. The internal validation exercise involved four holdout paired-comparison (PC) questions, which occurred immediately after the sixteen PC questions designed by the respective conjoint methods. The external validation exercise was the selection of a laptop computer bag from among five bags offered in a pareto set from which they could buy and keep the product (and cash). This exercise occurred in the same session as the conjoint tasks and holdout questions, but was separated from these activities by a filler task.[2]

[2] One month later the subjects were recontacted and given the opportunity to select features with a "configurator" that mimicked websites such as dell.com in which consumers select features from a menu. The analysis of this e-commerce task is beyond the scope of the present paper. However, the basic results followed the same pattern as the internal and external validity tests in this paper. For example, feature hit-rates were 0.77 for FP, 0.71 for Fixed, and 0.69 for ACA with the former significantly better than the latter two (0.05 level).

Table 1. Overview of the Research Design

FastPace	Fixed	ACA
FP Conjoint	Fixed Conjoint	ACA Conjoint
Holdout Questions	Holdout Questions	Holdout Questions
Choose and keep	Choose and keep	Choose and keep

Conjoint Tasks

Recall that ACA requires four sets of questions. In this case, pretests confirmed that all of our features were acceptable to the target market, allowing us to skip the unacceptability task. This left three remaining tasks: self-explicated (SE) questions, metric paired-comparison (PC) questions, and purchase intention (PI) questions. ACA uses the SE questions to select the PC questions, thus the SE questions in ACA must come first, followed by the PC questions and then the PI questions. To test ACA fairly, we adopted this question order for our ACA cell.

Although some hybrid conjoint methods exist in which SE questions are collected for each respondent and updated with an across-respondent efficient experimental design, the focus of the benchmark in this paper is on data collected to estimate partworths for each respondent. Because asking the SE questions first could create a question-order effect that confounds the comparison of FP with an efficient design of PC questions (and corresponding estimation methods), we asked only the PC questions (not the SE or PI questions) prior to the validation task in the fixed-efficient-design cell.

FP is designed to obtain estimates with as few questions as possible. These estimates are feasible based on the PC questions alone. However, Toubia, Simester and Hauser (2001) also explore FP hybrids that take advantage of SE questions. Thus, to avoid confounding question order with comparisons between purebred methods, we included two FP data collection procedures: one that matched that of ACA and one that matched that of fixed-efficient designs. In FP1 the PC questions precede the validation task. In FP2, the SE, PC, and PI questions precede the validation task. Not only does this allow a “clean” comparison of FP vs. ACA and of FP vs. Fixed, it enables us to explore whether or not answering the SE questions affects how respondents answer the PC questions, an effect documented in the literature (e.g., Alreck and Settle 1995; Green, Krieger and Agarwal 1991; Huber, et. al. 1993; Johnson 1991). For more general discussions of question-order effects, see Bickart 1993; Feldman and Lynch 1988; Nowlis and Simonson 1997; Tourangeau, Rips and Rasinski 2000.

Our complete research design, including the question order, is summarized in Table 2. Questions associated with the conjoint tasks are highlighted in green, while the validation tasks are highlighted in yellow. The filler task is highlighted in blue. In this design, FP1 can be matched with Fixed; FP2 can be matched with ACA.

Table 2. Detailed Research Design

FP1	Fixed	ACA	FP2
		Self-explicated	Self-explicated
FastPace paired comparison	Fixed paired comparison	ACA paired comparison	FastPace paired comparison
Holdout paired comparison	Holdout paired comparison	Holdout paired comparison	Holdout paired comparison
		Purchase intentions	Purchase intentions
Filler task	Filler task	Filler task	Filler task
Choose and keep product and/or cash	Choose and keep product and/or cash	Choose and keep product and/or cash	Choose and keep product and/or money

Holdout PC Questions

We allowed each of the conjoint methods to design sixteen metric paired-comparison (PC) questions. Following these sixteen questions, respondents were presented with four additional PC questions. These four questions served as holdout questions and were used to evaluate internal validity. This is a common test that has been used extensively in the literature.[3]

[3] The number of holdout questions varies, but is generally between two and four questions. Examples include Acito and Jain (1980), Akaah and Korgaonkar (1983), Cattin, Hermet and Pioche (1982), Elrod, Louviere and Davey (1992), Green, Goldberg and Wiley (1982), Green, Goldberg and Montemayor (1981), Green and Helsen (1989), Green, Helsen and Shandler (1988), Green and Krieger (1995), Green, Krieger and Agarwal (1991), Green, Krieger and Bansal (1988), Haaijer, Wedel, Vriens and Wansbeek (1998), Hagerty (1985), Huber (1975), Huber, Wittink, Fiedler and Miller (1993), Hauser and Koppelman (1979), Hauser and Urban (1977), Hauser, Tybout and Koppel-

The same procedure was used to design these four questions in each of the experimental conditions, and this design was independent of the sixteen questions that the respondents had just answered. In particular, the product profiles in these four questions were randomly selected from an independent efficient design of 16 profiles. There was no separation between the sixteen initial questions and the four holdout questions, so that respondents were not aware that the questions were serving a different role.

Filler Task

The filler task was designed to separate the conjoint tasks and the simulated store. It was hoped that this separation would mitigate any memory effects, which might influence how accurately the information from the conjoint tasks predicted the purchasing decisions in the respondents chose and kept the products and money. The filler task was the same in all four experimental conditions and comprised a series of questions asking respondents about their satisfaction with the survey questions. There was no significant difference in responses across the four conditions.

Choose and keep product and/or money (Simulated Store)

Respondents were told that they had $100 to spend and were asked to choose between five bags. The five bags shown to each respondent were randomly drawn from an orthogonal fractional factorial design of sixteen bags. This design was the same across all four experimental conditions, so that there was no difference on average in the bags shown to each respondent. The five bags were also independent of responses to the earlier conjoint questions. The price of the bags varied between $70 and $100 reflecting the difference in the market price of the features included with each bag. By pricing the bags in this manner we assured that the choice set represented a Pareto frontier as recommended by Elrod, Louviere, and Davey (1992), Green, Helsen and Shandler (1988), and Johnson, Meyer and Ghosh (1989).

Respondents were instructed that they would receive the bag that they chose. If the bag was priced at less that $100, they were promised cash for the difference. In order to obtain a complete ranking, we told respondents that if one or more alternatives were unavailable, they might receive a lower ranked bag. The page used to solicit these rankings is presented in Figure

---

man (1981), Jain, Acito, Malhotra and Mahajan (1979), Johnson, Meyer and Ghosh (1989), Johnson (1999), Lenk, DeSarbo, Green and Young (1996), Malhotra (1986), Moore (1980), Moore and Semenik (1988), Orme, Alpert and Christensen (1998), Orme and King (1998), Parker and Srinivasan (1976), and Tybout and Hauser (1981).

2. We acknowledge two tradeoffs in this design. First is an endowment effect because we endow each respondent with $100. The second is the lack of a “no bag” option. While both are interesting research opportunities and quite relevant to market forecasting, neither should favor one of the three methods relative to the other; the endowment/force-choice design is common to all treatments. Pragmatically, we selected an endowment/forced-choice task in order to maximize the statistical power with which to compare the four treatments.

Following our research, bags were distributed to respondents together with the cash difference (if any) between the price of the selected bag and $100.

Figure 2. Respondents Choose and Keep a Laptop Computer Bag

Features	first	fifth	second	third	fourth
Size	Large	Medium	Medium	Large	Large
Color	Red/Gray	Red/Gray	Black	Black	Red/Gray
Logo	Yes	Yes	Yes	Yes	No
Handle	No	No	Yes	Yes	Yes
PDA	No	No	No	No	No
Cell Phone	No	Yes	Yes	No	Yes
Mesh Pocket	Yes	No	Yes	No	No
Sleeve Closure	Full Flap	Tab Velcro	Full Flap	Tab Velcro	Tab Velcro
Boot	Yes	Yes	Yes	Yes	No
Price	$91	$79	$97	$87	$80

Self-Explicated and Purchase Intention Questions

The self-explicated questions asked respondents to rate the importance of each of the ten product features. For a fair comparison to ACA, we used the wording for the questions and (four point) response scale proposed by Sawtooth (Sawtooth 1996). Further details regarding the design of the questions and the measurement scales are available in Sawtooth (1996) and by looking on the website listed in the acknowledgements section of this paper (see also Figure 3d). For the purchase intentions questions respondents were shown six bags and were asked how likely they were to purchase each bag. We again adopted the wording, response scale, and algorithms for profile selection that are suggested by Sawtooth.[4]

Subjects

The subjects (respondents) were first-year MBA students. All had taken basic marketing, but none knew the details of our experiment and none had yet taken a course in which conjoint analysis is taught in detail. A total of 360 students were sent an email inviting their participation. We received 330 complete responses to the first stage of the study (there was one incomplete response) resulting in a response rate of over 91%. This high response rate reflects the perceived desirability of the bag relative to the short time investment required to respond. Random assignment yielded 80 subjects for the ACA condition, 88 for the Fixed condition, and 162 for the FP conditions broken out as 88 for the standard question order (FP1) and 74 for the alternative question order (FP2).

The questionnaires were pretested on a total of 69 subjects drawn from professional market research and consulting firms, former students, and students in an advanced marketing course that studied conjoint analysis. The pretests were valuable for fine-tuning the question wording and the web-based interfaces. By the end of the pretest, respondents found the questions unambiguous and easy to answer. Following standard scientific procedures, the pretest data were not merged with the experimental data. However, analysis of this small sample suggests that the findings agree directionally with those reported here, albeit not at the same level of significance. To ensure that the pretest subjects had similar incentives to the respondents in the actual study, pretest subjects were entered into a lottery, the winners of which received their chosen bag.

Additional Details

Figure 3 illustrates some of the key screens in the conjoint analysis questionnaires. In Figure 3a respondents are introduced to the price feature.[5] Figure 3b illustrates one of the dichotomous features – the closure on the sleeve. This is an animated screen which provides more

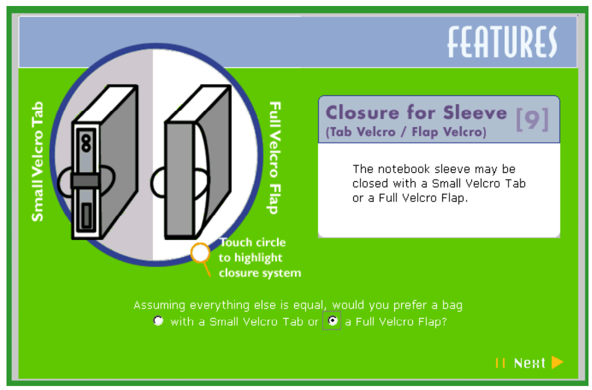
[4] The six profiles include the most attractive, the least attractive, and four “middling” profiles as determined by the respondent’s preferences. The four middling profiles are chosen to differ most strongly based on partworths as determined by the SE and PC questions.

[5] Due to human-subject concerns the details of the price-back-from-$100 incentive were described in an introductory screen which included other experimental protocols.

detail as respondents move their pointing devices past the picture. Figure 3c illustrates one of the PC tasks. The respondent was asked to choose between two profiles that varied on three features. Sawtooth (1996) recommends that respondents can handle three features once they are familiar with the task and cites experience that little benefit is gained by using more than three features. To provide a fair comparison to ACA, we adopted this guideline. Profiles were described by both text and pictures in which the non-varying features were chosen to coincide with the respondent's choices in the tasks illustrated in Figure 3b. The format was identical for all four experimental treatments. Finally, Figure 3d illustrates the first three self-explicated questions. The full questionnaires for each treatment are available at the website listed in the acknowledgements to this paper.

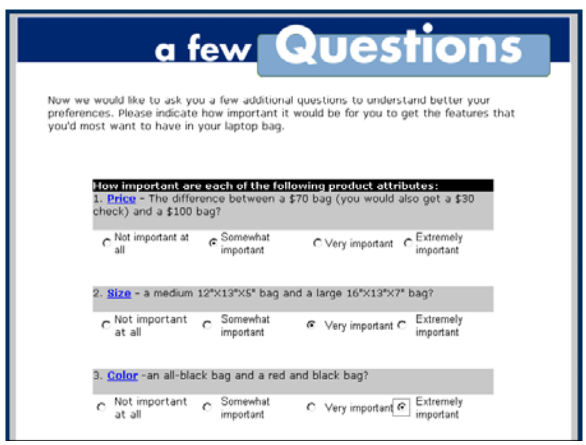
Figure 3. Example Screens from Questionnaires

(a) Price as change from $100



(b) Introduction of "sleeve" feature

(c) Metric paired-comparison question



(d) Self-explicated questions

Test of Internal Validity

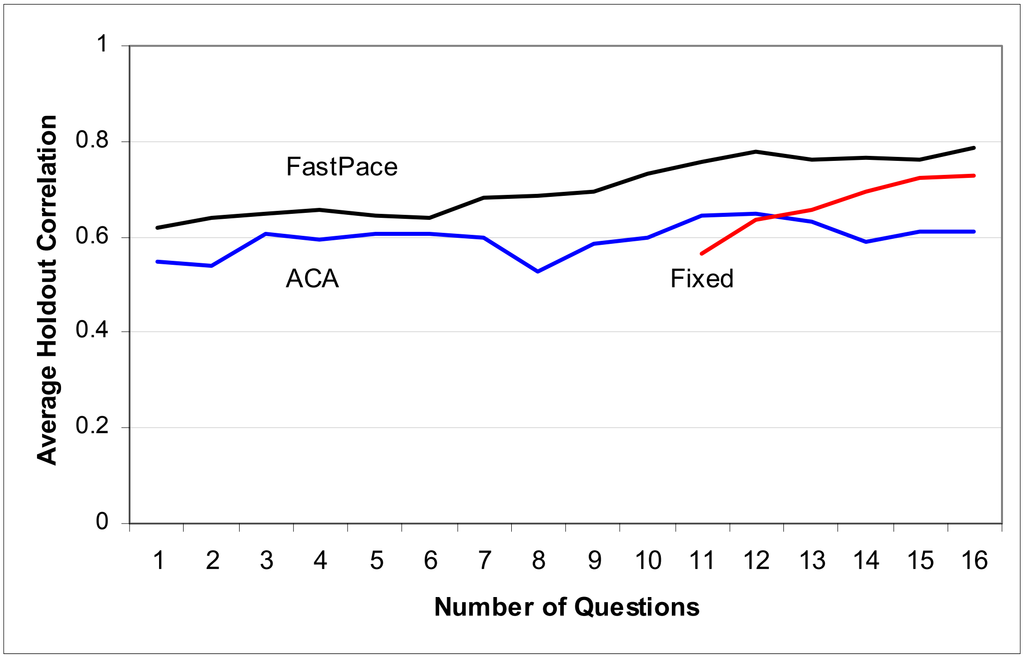
To test internal validity we compare the ability of each method to predict preferences in four holdout paired-comparison (PC) questions. For each respondent we calculated the correlation between the predicted and observed responses across the four questions. This correlation, averaged across respondents in each condition, is summarized in Figure 4. The correlations are recalculated after each additional PC question, so that we can observe the incremental accuracy when asking additional PC questions.

As an alternative metric, we also compared how well the methods predicted which product the respondents favored from each pair. The two metrics provide a very similar pattern of results and so for ease of exposition we focus on the correlation measure in Figure 4.[6] Because our initial tests focus on the comparison of the "purebred" methods, we use FP1 as representing FP. As it turns out, we obtain the same relative comparisons to ACA and Fixed when we use FP2. Thus, we delay the formal comparison between FP1 and FP2 to a later section.

The correlations are within the ranges reported in prior research. When metric data are gathered, correlations of predicted and observed ratings vary from 0.25 to 0.82 and are typically in the range of 0.60 to 0.75. For this product category and for this feature set, the adaptive methods, FP and ACA, appear to provide reasonable accuracy with far fewer questions than a non-adaptive method. We examine later whether this remains true for Hierarchical Bayes estimation.

[6] There is an additional reason to focus on correlations in the external validity tests. First preference prediction is a dichotomous variable and has higher variance than the Spearman correlation which is based on five observations per respondent.

Figure 4. Interval Validity Test – Holdout Pairs



For the fixed method, OLS estimation is only possible after at least eleven PC responses (there are ten parameters to estimate). Although accuracy after eleven questions is lower than the other two methods, accuracy improves when additional PC responses are received. Toubia, Simester and Hauser (2001) report simulations suggesting a similar pattern of results using a related fixed model. To give efficient fixed designs the best chance in their simulations, they addressed the situation where the researcher could choose the number of questions ahead of time for each number of questions (q=1 to 16) and optimize the design for each number of questions. Of necessity we designed a single set of sixteen questions for every respondent (randomizing the question order across respondents) and gathered data only up to $q=16$. Thus, their simulated fixed design provides an upper bound on the predictive ability of fixed designs. Despite these differences they found that fixed designs do not do as well as FP and ACA for small number of questions, but improve as the number of questions exceeds 1.5 times the number of parameters – 15 questions in our experiment. Were we to ask 20 questions from an efficient design, their simulations suggest that Fixed would approach the accuracy of FP in Figure 4.

To evaluate whether the difference in the performance of the three methods is statistically significant we first focused on the final estimates using responses to all sixteen PC questions. The analysis confirmed that the FP estimates are significantly ($p<0.05$) more accurate than the

ACA estimates. After sixteen questions, the Fixed method is also significantly ($p<0.05$) more accurate than the ACA method, but is not significantly different from the FP method.

Because our managerial goal was to evaluate the methods for low numbers of questions, we also evaluated significance by a method that pools the correlation measures calculated after each additional PC question. This resulted in a total of sixteen observations for each respondent in the FP and ACA conditions, and six observations for respondents in the Fixed condition. To control for the resulting heteroscedasticity we need to control for a question effect. In addition, there might be heterogeneity across respondents in relative preferences for features, a heterogeneity that could affect the prediction task. To control for this potential heteroscedasticity we use as a control the correlation that we would have obtained were the respondent to use the same partworths for all features (equal-weight model). With these controls we used OLS to estimate the following fixed effects model where *r* indexes respondent and *q* indexes the number of PC questions used in the partworth estimates. The α's and β's are coefficients in the regression and ε_{rq} is an error term:

(1) $$Correlation_{rq} = \alpha_0 + \alpha_1 Fixed + \alpha_2 ACA + \sum_{q=1}^{15} \beta_q Question_q + \beta_{ew} EW_r + \varepsilon_{rq}$$

The independent variables are defined as follows:

Fixed	1 if the respondent was in the Fixed condition; 0 otherwise.
ACA	1 if the respondent was in the ACA condition; 0 otherwise.
$Question_q$	1 if the correlation was calculated using only the first *q* PC questions; 0 otherwise.
EW_r	correlation obtained for respondent *r* with an equal-weights model

Under this specification, the intercept (α_0) describes the correlation for the FP condition, while the α_1 and α_2 coefficients represent the expected increase or decrease in this correlation in the Fixed and ACA conditions respectively. We also estimated a random effects model but there was almost no difference in the coefficients of interest. Moreover, the Hausman specification test favored the fixed-effects specification over the random-effects specification. The findings are presented in Table 3, where we also present findings from a similar analysis conducted on the external validity task. The β-coefficients for the variables describing the question and respondent heterogeneity are omitted for ease of exposition. We obtain similar results for holdout hit rates.

The findings from this multivariate analysis suggest a significant improvement ($p<0.05$) by the FP method relative to ACA and Fixed based on this internal validity task. However, there was no significant difference in the performance of the Fixed and ACA methods ($t=1.60$). We conclude that, based on the internal validity tests, the new polyhedral methods (FP) show promise, especially for obtaining reasonable estimates based on fewer questions. However, we now examine whether FP methods can be used to predict actual customer choice.

Table 3. Multivariate Analysis of the Pooled Data

	Holdout Paired Comparison Questions	Choose and Keep Product/Money
Intercept (α_0)	0.458* (0.027)	0.543* (0.026)
ACA (α_1)	-0.127* (0.014)	-0.142* (0.015)
Fixed (α_2)	-0.093* (0.021)	-0.288* (0.021)
Adjusted R^2	0.27	0.23
Sample Size	3,103	3,213

Coefficients describing the question and respondent heterogeneity are omitted from this table.
* Significant at $p<0.05$. Standard errors are in parentheses.

External Validity Test

While tests of internal validity are common in the conjoint-analysis literature, tests of external validity at the individual level are rare.[7] A search of the literature revealed four studies that predict choices in the context of natural experiments and one study based on a lottery choice. Wittink and Montgomery (1979), Srinivasan (1988), and Srinivasan and Park (1997) all use conjoint analysis to predict MBA job choice. Samples of 48, 45, and 96 (respectively) student subjects completed a conjoint questionnaire prior to selecting jobs. The methods were compared on their ability to predict those jobs. First preference predictions range from 64% to 76% versus random-choice percentages of 26-36%. This corresponds to an improvement relative to random in the range of 50-56%.[8]

In another natural experiment Wright and Kriewall (1980) used conjoint analysis (Linmap) to predict college applications by 120 families. They were able to correctly predict 20% of

[7] Some researchers report aggregate predictions relative to observed market share. See Bucklin and Srinivasan (1991), Currim (1981), Davidson (1973), Green and Srinivasan (1978), Griffin and Hauser (1993), Hauser and Gaskin (1984), McFadden (2000), Page and Rosenbaum (1989), and Robinson (1980).

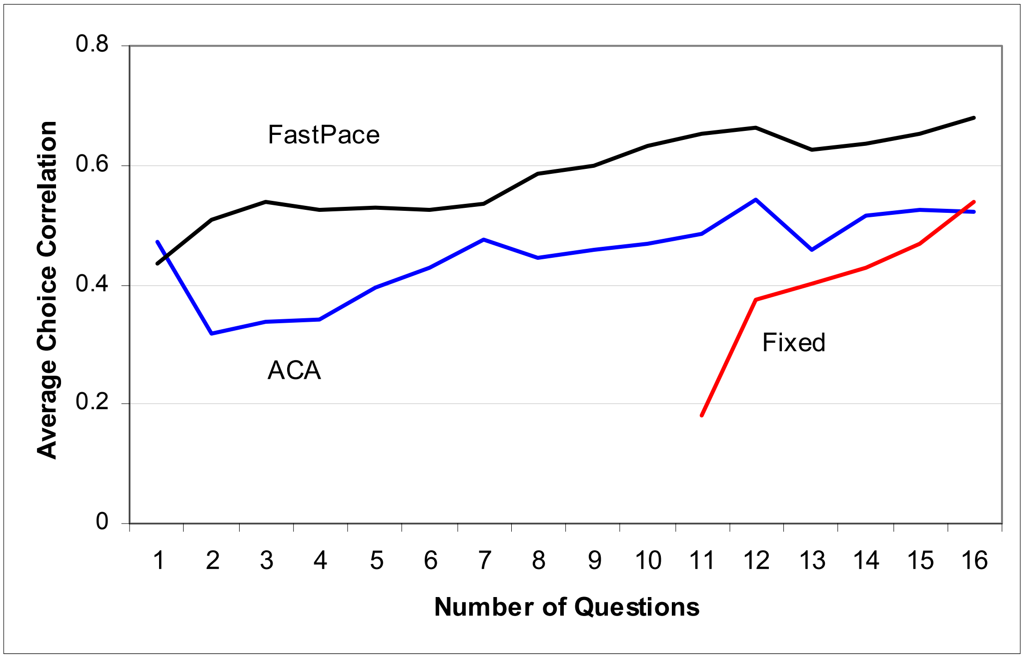
the applications when families were prompted to think seriously about the features measured in conjoint analysis; 15% when they were not. This converts to a 16% improvement relative to their null model. Leigh, MacKay and Summers (1984) allocated 122 undergraduate business majors randomly to twelve different conjoint tasks designed to measure partworths for five features. Respondents indicated their preferences for ten calculators offered in lottery. There were no significant differences among methods with first-preference predictions in the range of 26-41% and percentage improvements of 28%. The authors also compared the performance of estimates based solely on SE responses and observed similar performance to the conjoint methods.

In the related choice-based conjoint literature, Louviere, Hensher, and Swait (2000) compare the convergent validity between revealed preference and "stated preference" methods to suggest that, in most cases, the two methods measure the same relative coefficients (partworths). Although this is not predictive validity per se, it does provide evidence that stated preference questions, such as paired comparison (PC) questions, provide data consistent with actual choices made by consumers. Urban and Katz (1983) provide further evidence that models based on PC questions, such as their Assessor model, predict well.

Against this background we examine the results from the simulated store in which respondents were presented with a selection five bags, asked to select (and "purchase") one bag, and, in case their chosen bag was unavailable, to rank all five bags in terms of choice. To evaluate the conjoint methods we calculated the correlation between the actual and observed rankings for the five bags shown to each respondent. We then averaged these correlations across respondents in each condition. (Results based on first choice were qualitatively similar.) The results are presented in Figure 5, where we again present findings calculated after each of the PC questions.

[8] This commonly reported metric is calculated as (predicted – random)/(100% - random).

Figure 5. External Validity Test – Correlation with Actual Choice



The findings from this task are similar to the results of the internal validity test. The FP method yields more accurate predictions of the respondents' bag rankings than either the ACA or Fixed methods. When using all sixteen PC responses, the FP estimates are significantly more accurate than both the ACA estimates ($p<0.05$) and the Fixed estimates ($p<0.05$). The difference between the Fixed and ACA methods is not significant. We also compared the methods by re-estimating Equation 1 using the correlations from the simulated store data. The findings from this pooled data are reported in Table 3. With respect to FP, they reveal an similar pattern of results, with FP significantly ($p<0.05$) more accurate than ACA and Fixed. However, for this task ACA is significantly ($p<0.05$) more accurate than the Fixed method over the course of the questions for which they can be compared.

Inspection of Figure 5 reveals two other qualitative findings of note. First, the performance of the Fixed method is not as good relative to the other methods in this task as in the internal validity task. Second, ACA's predictive ability appears to first decrease as a function of q and then increase. Toubia, Simester and Hauser (2001) and Johnson (1987) report similar results when ACA's SE measures are accurate relative to the PC responses. The addition of noisy PC responses initially reduces overall accuracy because the information in these responses is outweighed by the noise that they introduce relative to the SE measures. Eventually, when there are sufficient PC responses, the information in these responses outweighs the noise.

It is helpful also to compare how frequently the different methods identified which bag each respondent most preferred. In the full information case (after all sixteen questions) FP correctly identified the first preference bag for approximately 59% of respondents, compared to 52% for the Fixed method and 50% for ACA.[9] A random selection would have chosen the first preference bag just 20% of the time. These correspond to percentage improvements of 46%, 40% and 38% for the FP, Fixed and ACA methods respectively. These percentage improvements are similar to the ranges reported elsewhere in the literature.

Summary of Validity Tests

Table 4 reviews the performance of the three methods on the internal and external validation metrics based on the sixteen questions. The findings are remarkably consistent across the two validity tasks. The estimates from the FP method are significantly more accurate than the ACA and Fixed estimates, suggesting that product-development teams should feel comfortable in using FP-based methods.

Recall that polyhedral methods for conjoint analysis were developed to provide reasonable estimates with fewer questions, especially for web-based questionnaires and for use in the new product development processes. In our tests, the FP's predictive abilities are relatively high for even a small number of questions, suggesting that the potential, subject to further testing, of using FP to screen features using relatively few questions. Moreover, given further development and refinement, it seems likely that researchers will develop polyhedral methods that can do even better than the current FP methods.

Although FP was significantly better than ACA and the fixed, non-adaptive design in our tests, neither ACA nor Fixed should be rejected. Both do well and both have proven valuable in many managerial applications. ACA and Fixed have strengths that might allow them to outshine FP in other empirical contexts. Because aspects of these methods might be combined with aspects of the FP method to produce an even better method, we next evaluate opportunities to improve predictions through hybrid techniques that draw features from each of these three methods.

[9] FP is significantly better than ACA ($p<0.05$) and not quite significant for Fixed ($p<0.07$).

Table 4. Summary of the Performance Metrics

	ACA	Fixed	FastPace
Holdout Paired Comparisons			
Correlation with holdout pairs *,+	0.61	0.73	0.79
Hit Rate for holdout pairs *, o, +	0.79	0.87	0.89
Choose and Keep Product/Money			
Correlation with choice among products *, o	0.52	0.54	0.68
First Preference Accuracy *	0.50	0.52	0.59

This summary is based predictions that use all sixteen PC questions.
* Significant difference ($p<0.05$) between FP and ACA.
o Significant difference ($p<0.05$) between FP and Fixed.
+ Significant difference ($p<0.05$) between Fixed and ACA.

Question-Order Effects – FP1 vs. FP2

ACA is based on a question order in which the self-explicated (SE) question precede the paired-comparison (PC) questions. Both FP and Fixed are based on the PC questions alone and do not require prior SE questions. The comparisons in the previous sections focused on the overall methods. However, it is possible that this comparison is due to question order. On hypothesis is that the SE questions tire respondents causing them to pay less attention to the paired-comparison (PC) questions (e.g., Alreck and Settle 1995). If this were true, then question order could explain the improved performance of FP (or Fixed) relative to ACA. A counter hypothesis is that the SE questions are a training task that improves the accuracy of the PC questions. For example, some researchers have observed a learning phenomenon in which the SE questions help respondents to clarify their values so as to increase the accuracy of the PC questions (Green, Krieger and Agarwal 1991; Huber, et. al. 1993; Johnson (1991). This effect might be task learning, self-preference learning, memory accessibility, or other context effects, but whatever the explanation, the "learning" effect might counter the wearout effect (Bickart 1993; Feldman and Lynch 1988; Nowlis and Simonson 1997; Tourangeau, Rips and Rasinski 2000; Simmons, Bickart, and Lynch 1993). Thus, the directional impact of question order is an empirical issue.[10]

[10] We might also observe a question-order effect if the PC questions preceded the SE questions or if the choice task preceded the SE questions. For example, Green, Krieger and Agarwal (1991) observe an effect that is consistent with PC-affects-SE hypothesis, but Huber, et. al. (1993) do not. In addition, theories in Allen 1982; Allen and Dillon 1982; Bem 1972; Folkes and Kiesler 1991; and Tybout and Yalch 1980 suggest that SEs that follow a choice task might be stated to justify choices.

ACA requires a specific question order, but FP does not. By comparing FP1, in which SEs does not precede PCs, and FP2, in which SEs precede PCs, we can examine whether or not there is a significant question-order effect. If the effect favors FP1 relative to FP2, then the we must use FP2 rather than FP1 to parse the improved performance of FP relative to ACA. If the effect favors FP2 relative to FP1, then our results are strengthened and FP2 becomes the better version of FP. In either case, if there is an effect, then future experiments can explore why there is an effect. On the other hand, if there is no effect, then our "purebred" comparisons survive.

We begin with the correlations and hit rates, both internal and external, based on all sixteen questions. Using the full sixteen questions, the predictive accuracy of the FP1 and FP2 conditions are not statistically different. This suggests that there was no measurable wearout due to the SE questions and that by the sixteenth question any warm-up/learning advantage had disappeared. However, there might still be an effect for low numbers of questions. If so, this would be important for the use of such estimates. Plotting FP1 and FP2 versus the number of questions for $q = 1$ to 16, we do observe that the correlations for FP2 are slightly higher than those for FP1 for low numbers of questions. However, when we use a version of Equation 1 to account for the heterogeneity of respondents, the effect is found to be not significant for the choice correlations ($t = 0.72$), the choice hit rate ($t = 0.34$), and the holdout hit rate ($t = 0.72$). FP2 is better on the holdout correlations ($t = 2.48$). Thus, only one of four multiple-question tests, does FP2 exceed FP1 and if we correct the α-level for multiple tests, this is not significant. Thus, for our data, there does not seem to be a strong question order effect and, if there is any effect at all, it favors FP2. Based on these tests we are more confident that the FP-vs.-ACA and Fixed-vs.-ACA comparisons in Figures 4-5 and Tables 3-4 are not due to question order and are likely due to other differences in the question-selection and/or estimation methods.

Improving Predictive Performance with Hybrid Methods

The term heterosis is used in biology to describe the principle under which genetically-diverse offspring of two pure species often have traits superior to either of their purebred parents (Campbell 1996, p. 428). Analogously, there is evidence that combinations of conjoint methods often yield more accurate or more efficient predictions than either of the parent methods.[11] Suc-

[11] See for example: Akaah and Korgaonkar 1983; Cattin, Hermet, and Pioche 1982; Carroll and Green 1995, Green 1984; Green, Goldberg and Montemayor 1981; Green, Goldberg and Wiley 1982; Green and Krieger 1995; Huber 1974; Huber, et. al. 1993; Moore and Semenik 1988; Wind, et. al. 1989.

cessful hybrid conjoint models have combined the power of compositional methods, such as self-explicated (SE) questions, and decompositional methods, such as paired-comparison (PC) questions to produce new estimates. In this context, ACA itself can be considered a hybrid. Although there are instances where purebred methods either outperform or provide equivalent accuracy to hybrid methods, there are many situations and product categories in which hybrid methods outperform their purebred progenitors (Green 1984; Wittink and Bergestuen 2001). Empirical tests suggest that both SE and PC questions add incremental information to one another (Green, Goldberg, and Montemayor 1981; Huber, et. al. 1993, Johnson 1999; Leigh, MacKay, and Summers 1984).

In our evaluation of hybrid methods we consider methods that combine either the question selection component or the estimation component of two (or more) methods to produce a new combined method. For example, we consider polyhedral methods that draw on the data contained in the SE responses.

Hybrid Estimation Procedures that Use the FastPace Question-Selection Method

Of the purebred methods only ACA uses the SE responses and Figures 4-5 suggest that the SE responses play a major role in the accuracy of ACA. Specifically, in our data there is little improvement in accuracy with ACA beyond that achieved with a single ACA-chosen PC question. Indeed, the PC questions initially degrade accuracy. This suggests that there is valuable information in the SE questions that might improve the FP predictions. In this section we hold the FP question-selection method constant and investigate the accuracy of alternative, hybrid estimation methods. Because we need the SE questions, we use the FP2 cell.

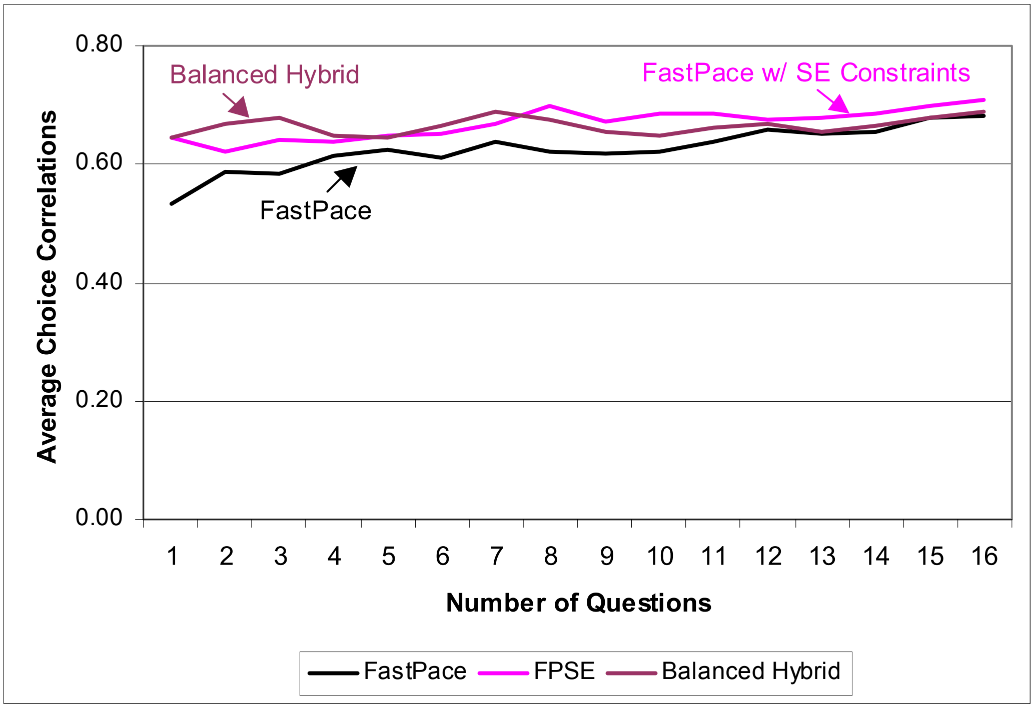
In our first benchmark, which we label *Balanced Hybrid*, we combine the SE and PC responses by minimizing the a least squares norm in which both the SE responses and the PC responses are treated as data. For p SE responses and q PC responses, the regression has $p+q$ observations that reflect identity for the SE responses and a linear utility difference for the PC responses. This procedure is similar to the approach suggested by Sawtooth (Sawtooth 1999) as a modification of the ACA method and is similar to the OLS model that ACA uses to obtain intermediate partworth estimates.

In second benchmark, which we label *FastPace with SE-based Constraints (FPSE)*, we introduce the SE responses as constraints on the feasible polyhedron that forms the basis of the FP estimation procedure. For example, if Feature 1 is given a higher SE importance weight than

Feature 2, we add a constraint to the FP polyhedron such that the partworth of Feature 1 is larger than the partworth of Feature 2.

Figure 6 compares the two hybrid methods to a purebred FP (repeating the FP data from Figure 5). Because we obtain similar results for both the holdout pairs and the choice of bags, we report only the latter correlations in Figure 6. Based on Equation 1, both hybrids are significantly better ($p<0.05$) that a purebred FP. Of the two hybrids, FPSE appears to do slightly better than the Balanced Hybrid. However, this difference is not significant. Thus, it appears that both the PC and the SE questions add incremental information and that hybrid methods have the potential to predict significantly better than either FP or SE alone – at least when the PC questions are chosen by FP.[12]

Figure 6. Hybrid Methods Based on FP Question-Selection



These findings are consistent with Toubia, Simester and Hauser's (2001) simulations which suggest that SE-PC hybrids have the potential to improve FP. These findings also suggest that further research on hybrid estimation could improve the methods further. However, we must caution the reader that these estimation procedures were tested <u>after</u> the data were collected.

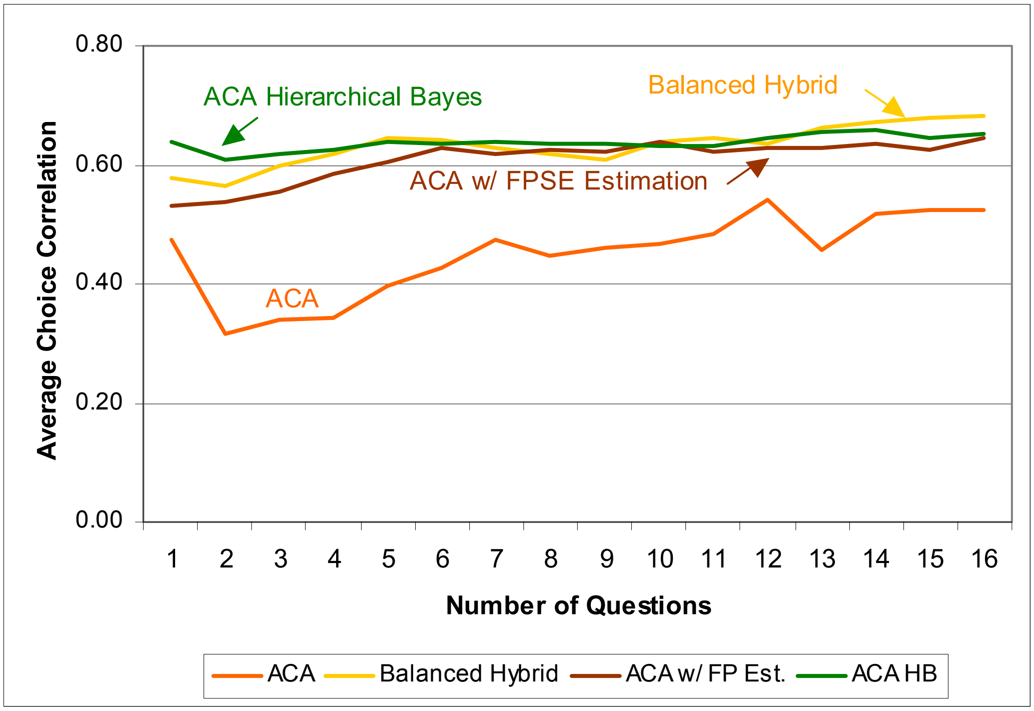
[12] For q=16 on the FP2 sample, both hybrids and FP perform better than SE alone. Average correlations are 0.71, 0.69, 0.68, and 0.64 for FPSE, Balanced Hybrid, FP, and SE, respectively.

Alternative Estimation Procedures that Use the ACA Question-Selection Method

For ACA question selection we estimate the same hybrid models as we did for FP question selection. That is, we apply FPSE estimation and Balanced Hybrid estimation using the PC questions as selected by the ACA algorithm. In addition we attempt to improve predictions with Hierarchical Bayes (HB) estimation. HB is based on recent developments in the estimation of conjoint-analysis data which suggest that predictions at the level of the individual respondent can be improved by using data from the population of respondents. These methods use the data from the full sample to iteratively estimate individual-level partworths and do so by constraining the distribution of those partworths. See Allenby and Rossi 1999; Arora, Allenby, and Ginter 1998; Lenk, et. al. 1996; Liechty, Ramaswamy and Cohen 2001. The estimation is based on Gibbs sampling and the Metropolis Hastings Algorithm (Sawtooth 1999). Sawtooth's Hierarchical Bayes algorithm allows a number of ways in which the SE data are used. For a fair comparison with purebred ACA, we chose the method recommended by Sawtooth which uses the SE responses to provide initial approximations.

The predictions based on these hybrid methods are shown in Figure 7. Based on the multiple-question statistical tests (Equation 1), all three hybrid methods are significantly better ($p<0.05$) than the purebred ACA. Thus, for ACA-based questions, it appears that here, too, heterosis improves predictions. ACA HB is particularly interesting in Figure 7. By combining population-level data with accurate individual-level SE data, it does quite well, even for a single question. Interestingly, as the number of PC questions increases and, hence, the number of observations per respondent increases, the Balanced Hybrid method, which relies only on data from each respondent, appears to outperform ACA HB. However, this improvement is not significant.

Figure 7. Hybrid Methods Based on ACA Question-Selection



Hybrid Estimation Procedures that Use the Fixed Question-Selection Method

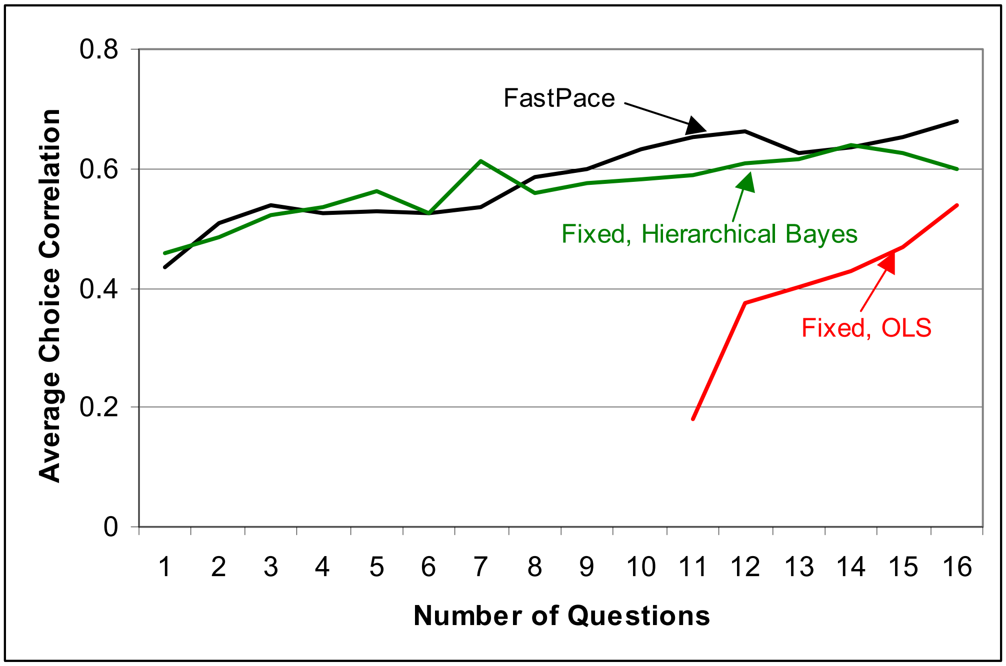
In the Fixed-efficient-design cell the only preference data available are the PC questions. Nonetheless, we can improve the predictive ability of these questions as well. In particular, the Hierarchical Bayes (HB) methods combine population-level data with paired comparison data from each individual to obtain estimates of partworths for each individual.[13] To understand the method intuitively, consider the special case when the design matrix is full rank (as is our matrix for $q = 16$). In this case, the HB estimate is a weighted average of the OLS estimator for each respondent and the pooled population estimate where the tradeoff is based on a complex function of the estimated variance structure (Lenk, et. al. 1996, p. 177). Thus, HB provides an alternative method by which we might estimate individual-respondent-level partworths with a small number of questions. For example, Lenk, et. al. (1996) report internal validity tests in which the root-mean-squared errors with HB for $q = 4$ are comparable to those obtained with OLS for $q = 16$.

Figure 8 reports the results obtained for our external validity data for an HB method based on the PC questions alone. For comparison, we repeat the data for the FP and Fixed meth-

[13] Technically, the Hierarchical Bayes method estimates posterior distributions of those partworths. For ease of comparison, we work here with the mean of those distributions for each respondent.

ods. Based on the Equation 1 tests, HB is significantly better than OLS for the Fixed sample ($p < 0.05$). On the other hand, FP is (is not) significantly better than HB.

Figure 8. Hierarchical Bayes Improvement for Efficient Fixed Question Selection



Summary of Hybrid Analyses

While we are encouraged by these findings, we must be cautious in our interpretations. The tested methods do well and all add significantly to a product development teams' knowledge of customer preference. The SE-based hybrids appear to improve predictions, but are based on post analysis of the data in a category where SE questions do well. It should not surprise us that the purebred methods can be improved with such hybrids in this category. (We chose our category to be conservative in our test of FP versus ACA, but any choice that favors ACA also favors hybrids.) Because the accuracy of SE responses varies, SE-based hybrids may not do well in all categories (e.g., Akaah and Korgaonkar 1983; Green 1984; Green, Goldberg and Wiley 1982; Green, Krieger and Agarwal 1991; Sheth and Talarzyk 1972). Before embracing SE-based hybrids for all categories, we must await further testing in categories in which SE responses are less accurate. Such categories include situations where the product is less separable and more holistic and categories where the direct scaling of partworths by respondents is difficult. We expect that such categories exist. For example, in a review of forty-two SE applications, Wilkie and Pessemier (1973) identify 45% with favorable results, 33% questioning the basic model, and the remainder mixed.

HB estimation also appears to do well and provides an alternative method to obtain individual-respondent-level estimates with fewer questions than there are parameters. Consistent with the literature, there is no question that HB improves predictions relative to OLS (cf. Judge, et. al., 1985, Chapter 3), but it remains an open question whether HB methods have superior external validity to purebred FP (Figure 8) or to ACA Balanced Hybrid (Figure 7). Hopefully, this issue can be researched further as all three methods improve.

It is tempting to compare across Figures 6, 7, and 8 to get some insight on the question-selection method. For example, the best method based on FP-chosen paired-comparison questions, FPSE, appears to predict better than the best estimation methods, Balanced Hybrid and Hierarchical Bayes, based on ACA-chosen paired-comparison questions, and the best method, Hierarchical Bayes, based on a fixed efficient design. However, controlling for respondent heterogeneity with Equation 1, this comparison is (is not) significant.

Discussion

Polyhedral methods show promise as methods to improve question selection and/or partworth estimation in conjoint analysis. Such methods are particularly promising for use as product-development processes become more iterative, dispersed, and global. These methods provide capabilities for product development teams to screen large numbers of features with relatively few questions per respondent. This advantage is also valuable for web-based interviewing if respondents are sensitive to having to answer too many PC questions. Polyhedral methods do well in simulation, but, prior to this paper, they were unproven in empirical situations.

This paper provides initial tests of one polyhedral method with actual web-based respondents in a real choice situation for a product that was new to the market at the time of the test. Our tests include the traditional internal validity (holdout) test as well as an external validity test – the actual choice from a set of five laptop computer bags. The internal and external tests are consistent with one another and consistent with prior simulations. Most importantly, all tested preference measurement methods do well and add incremental information relative to an equal weights model. Both the self-explicated (SE) and the metric paired-comparison (PC) questions appear to add incremental information to one another as is consistent with the literature. In this first head-to-head comparison, the FastPace (FP) method appears to improve predictions relative to the benchmarks of ACA and (fixed) efficient experimental designs.

We also tested several hybrid methods. In our data, the hybrids do well and improve predictions relative to the purebred methods. While this may be unique to situations where self-explicated questions are highly accurate, these tests suggest that hybrid improvements are worth considering in future empirical situations. The greatest improvement for the FP method appears to be the use of self-explicated responses to impose constraints on the feasible polyhedron – a method we have labeled FPSE. The greatest improvement for ACA appears to be a hybrid estimation method (from a previous version of ACA) in which data from the SE and PC questions are balanced. Hierarchical Bayes (HB) estimation also improves ACA with slightly better prediction than the balanced method for low numbers of PC questions and slightly lower prediction for high numbers of PC questions. HB also does well for the fixed, efficient experiment designs providing an alternative method for obtaining partworth estimates with fewer questions than there are parameters.

Our research design also enabled us to examine question-order effects. Specifically, although the prior SE questions in FP2 can improve predictions with hybrids, they do not appear to either enhance or degrade the information in the PC questions.

In the end we conclude that all of the methods tested appear to perform well on these validity tests. This is good news for the many academics and product-development professionals that rely on conjoint analysis. We conclude further that the polyhedral methods appear to have passed their first empirical test. The performance of FP and the FP Hybrids suggest that further developments based on polyhedral methods have the potential to advance conjoint analysis theory and practice. We hope that by making our methods, our source code, and our data public, we can encourage such innovation.

References

Acito, Franklin and Arun K. Jain (1980), “Evaluation of Conjoint Analysis Results: A Comparison of Methods,” *Journal of Marketing Research*, 17, (February), 106-112.

Akaah, Ishmael P. and Pradeep K. Korgaonkar (1983), “An Empirical Comparison of the Predictive Validity of Self-explicated, Huber-hybrid, Traditional Conjoint, and Hybrid Conjoint Models,” *Journal of Marketing Research*, 20, (May), 187-197.

Allen, Chris T. (1982), “Self-Perception Based Strategies for Stimulating Energy Conservation,” *Journal of Consumer Research*, 8 (March), 381-390.

_____, and William R. Dillon (1983), “Self-Perception Development and Consumer Choice Criteria: Is There A Linkage?” in *Advances in Consumer Research*, 10, Eds. Richard P. Bagozzi and Alice M. Tybout. Ann Arbor, MI: Association for Consumer Research, 45-50.

Allenby, Greg M. and Peter E. Rossi (1999), “Marketing Models of Consumer Heterogeneity,” *Journal of Econometrics*, 89, (March/April), 57-78.

Alreck, Pamela L. and Robert B. Settle (1995), The Survey Research Handbook: Guidelines and Strategies for Conducting a Survey, 2E, (New York, NY: McGraw Hill), p. 102.

Arora, Neeraj, Greg M. Allenby and James L. Ginter (1998), “A Hierarchical Bayes Model of Primary and Secondary Demand,” *Marketing Science*, 17, 1, 29-44.

Ben-Akiva, Moshe and S. Gershenfeld (1998), “Multi-featured Products and Services: Analyzing Pricing and Bundling Strategies,” *Journal of Forecasting*, 17, 175-196.

Bem, Daryl (1972), “Self Perception Theory,” in *Advances in Experimental Social Psychology*, 6, Ed. Leonard Berkowitz, New York: Academic Press.

Bickart, Barbara A. (1993), “Carryover and Backfire Effects in Marketing Research,” *Journal of Marketing Research*, 30, 1, (February), 52-62.

Bucklin, Randolph E. and V. Srinivasan (1991), “Determining Interbrand Substitutability Through Survey Measurement of Consumer Preference Structures,” *Journal of Marketing Research*, 28, (February), 58-71.

Campbell, Neil A. (1996), *Biology, 4E*, (Reading, MA: The Benjamin/Cummings Publishing Company, Inc.).

Carroll, J. Douglas and Paul E. Green (1995), “Psychometric Methods in Marketing Research: Part I, Conjoint Analysis,” *Journal of Marketing Research*, 32, (November), 385-391.

Cattin, Philippe and Dick R. Wittink (1982), “Commercial Use of Conjoint Analysis: A Survey,” *Journal of Marketing*, 46, (Summer), 44-53.

______, Garard Hermet and Alain Pioche (1982), “Alternative Hybrid Models for Conjoint Analysis: Some Empirical Results,” in *Analytic Approaches to Product and Marketing Planning: The Second Conference*, R. K. Srinivasan and Allen D. Shocker, eds., (Cambridge, MA: Marketing Science Institute), 142-152.

Choi, S. Chan and Wayne S. DeSarbo (1994), “A Conjoint-based Product Designing Procedure Incorporating Price Competition,” *Journal of Product Innovation Management*, 11, 451-459.

Currim, Imran S. (1981), “Using Segmentation Approaches for Better Prediction and Understanding from Consumer Mode Choice Models,” *Journal of Marketing Research,* 18, (August), 301-309.

Cusumano, Michael A. and Richard W. Selby (1995), *Microsoft Secrets*, (New York, NY: The Free Press).

Dahan, Ely and John R. Hauser (2000), "The Virtual Customer," working paper, Cambridge, MA: Center for Innovation in Product Development, MIT, (June), forthcoming, *Journal of Product Innovation Management*.

______ and V. Srinivasan (2000), “The Predictive Power of Internet-Based Product Concept Testing Using Visual Depiction and Animation,” *Journal of Product Innovation Management*, 17, 99-109.

Davidson, J. D. (1973), “Forecasting Traffic on STOL,” *Operations Research Quarterly*, 22, 561-569.

Dawes, Robin M. and Bernard Corrigan (1974), “Linear Models in Decision Making,” *Psychological Bulletin*, 81, (March), 95-106.

De Angelis, Chris (2001), “Sampling for Web-Based Surveys,” paper presented at 22nd Annual Marketing Research Conference, Atlanta, GA, September.

Einhorn, Hillel J. (1971), “Use of Nonlinear, Noncompensatory, Models as a Function of Task and Amount of Information,” *Organizational Behavior and Human Performance*, 6, 1-27.

Elrod, Terry, Jordan Louviere, and Krishnakumar S. Davey (1992), “An Empirical Comparison of Ratings-Based and Choice-based Conjoint Models,” *Journal of Marketing Research* 29, 3, (August), 368-377.

Eppinger, Steven D. (1998), “Information-Based Product Development,” Presentation to the Research Advisory Committee of the Center for Innovation in Product Development, MIT, Cambridge, MA (January 12).

_______, Daniel E. Whitney, Robert P. Smith, and David A. Gebala (1994), "A Model-Based Method for Organizing Tasks in Product Development", *Research in Engineering Design*. 6, 1, 1-13.

Feldman, Jack M. and John G. Lynch, Jr. (1988), “Self-generated Validity: Effects of Measurement on Belief, Attitude, Intention, and Behavior,” *Journal of Applied Psychology*, 73, (August), 421-435.

Folkes, Valerie S., and Tina Kiesler, (1991), “Social Cognition: Consumers’ Inferences About the Self and Others,” in Thomas S. Robertson and Harold H. Kassarjian, eds., *Handbook of Consumer Behavior*, 281-315.

Freund, Robert (1993), “Projective Transformations for Interior-Point Algorithms, and a Superlinearly Convergent Algorithm for the W-Center Problem,” *Mathematical Programming*, 58, 385-414.

Gonier, Dennis E. (1999), “The Emperor Gets New Clothes,” Paper presented at the Advertising Research Foundation’s On-line Research Day and available at www.dmsdallas.com. (January).

Green, Paul E., (1984), “Hybrid Models for Conjoint Analysis: An Expository Review,” *Journal of Marketing Research*, pp. 155-169.

Green, Paul E. and Kristiaan Helsen (1989), “Cross-Validation Assessment of Alternatives to Individual-Level Conjoint Analysis: A Case Study,” *Journal of Marketing Research*, pp. 346-350.

______, ______, and Bruce Shandler (1988), “Conjoint Internal Validity Under Alternative Profile Presentations,” *Journal of Consumer Research*, 15, (December), 392-397.

______, Stephen M. Goldberg, and Mila Montemayor (1981), “A Hybrid Utility Estimation Model for Conjoint Analysis,” *Journal of Marketing*, pp. 33-41.

______, ______, and James B. Wiley (1982), “A Cross Validation Test of Hybrid Conjoint Models,” in *Advances in Consumer Research*, 10, Richard P. Bagozzi and Alice M. Tybout, eds., (Ann Arbor, MI: Association for Consumer Research), 147-150.

______ and Abba Krieger (1995), “Attribute Importance Weights Modification in Assessing a Brand’s Competitive Potential,” *Marketing Science*, 14, 3, Part 1 of 2, 253-270.

______, ______, and Manoj K. Agarwal (1991), “Adaptive Conjoint Analysis: Some Caveats and Suggestions,” *Journal of Marketing Research*, pp. 215-222.

______, ______, and Pradeep Bansal (1988), “Completely Unacceptable Levels in Conjoint Analysis: A Cautionary Note,” *Journal of Marketing Research*, 25, (August), 293-300.

______ and V. Srinivasan (1990), “Conjoint Analysis in Marketing: New Developments With Implications for Research and Practice,” *Journal of Marketing*, pp. 3-19.

______ and ______ (1978), “Conjoint Analysis in Consumer Research: Issues and Outlook,” *Journal of Consumer Research*, 5, 2, (September), 103-123.

Griffin, Abbie and John R. Hauser (1993), "The Voice of the Customer," *Marketing Science*, vol. 12, No. 1, (Winter), 1-27.

Haaijer, Rinus, Michel Wedel, Marco Vriens, and Tom Wansbeek (1998), “Utility Covariances and Context Effects in Conjoint MNP Models,” *Marketing Science*, 17, 3, 236-252.

Hagerty, Michael R. (1985), “Improving the Predictive Power of Conjoint Analysis” The Use of Factor Analysis and Cluster Analysis,” *Journal of Marketing Research*, 22, (May), 168-184.

Hauser, John R., and Steven P. Gaskin (1984), "Application of the `Defender' Consumer Model," *Marketing Science*, Vol. 3, No. 4, (Fall), 327-351.

______ and Frank S. Koppelman (1979), "Alternative Perceptual Mapping Techniques: Relative Accuracy and Usefulness, *Journal of Marketing Research*, Vol. 16, No. 4, (November), 495-506.

______ and Steven M. Shugan (1980), “Intensity Measures of Consumer Preference,” *Operations Research*, 28, 2, (March-April), 278-320.

______, Alice M. Tybout, and Frank Koppelman (1981), "Consumer-Oriented Transportation Service Planning: "Consumer Analysis and Strategies," *Applications of Management Science*, Vol. 1, 91-138.

______ and Glen L. Urban (1977), “A Normative Methodology for Modeling Consumer Response to Innovation,” *Operations Research,* 25, 5, (July-August), 579-619.

Huber, George P. (1974), “Multiattribute Utility Models: A Review of Field and Field-like Studies, *Management Science*, 20, (June), 1393-1402.

Huber, Joel (1975), “Predicting Preferences on Experimental bundles of Attributes: A Comparison of Models,” *Journal of Marketing Research*, 12, (August), 290-297.

______, Dick R. Wittink, John A. Fiedler, and Richard Miller (1993), “The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice,” *Journal of Marketing Research*, pp. 105-114.

Jain, Arun K., Franklin Acito, Naresh K. Malhotra, and Vijay Mahajan (1979), "A Comparison of the Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models," *Journal of Marketing Research*, 16, (August), 313-322.

Johnson, Eric J., Robert J. Meyer, and Sanjoy Ghose (1989), "When Choice Models Fail: Compensatory Models in Negatively Correlated Environments," *Journal of Marketing Research*, pp. 255-270.

Johnson, Richard (1987), "Accuracy of Utility Estimation in ACA," Working Paper, Sawtooth Software, Sequim, WA, (April).

______ (1991), "Comment on `Adaptive Conjoint Analysis: Some Caveats and Suggestions," *Journal of Marketing Research*, 28, (May), 223-225.

______ (1999), "The Joys and Sorrows of Implementing HB Methods for Conjoint Analysis," Working Paper, Sawtooth Software, Sequim, WA, (November).

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T. C. Lee (1985), *The Theory and Practice of Econometrics*, (New York, NY: John Wiley and Sons).

Karmarkar, N. (1984), "A New Polynomial Time Algorithm for Linear Programming," *Combinatorica*, 4, 373-395.

Keeney, Ralph and Howard Raiffa (1976), *Decisions with Multiple Consequences: Preferences and Value Tradeoffs*, (New York, NY: John Wiley & Sons).

Klein, Noreen M., "Assessing Unacceptable Attribute Levels in Conjoint Analysis," *Advances in Consumer Research* vol. XIV, pp. 154-158.

Krantz, David H., R. Duncan Luce, Patrick Suppes, and Amos Tversky (1971), *Foundations of Measurement*, (New York, NY: Academic Press).

Kuhfeld, Warren F. (1999), "Efficient Experimental Designs Using Computerized Searches," Working Paper, SAS Institute, Inc.

______, Randall D. Tobias, and Mark Garratt (1994), "Efficient Experimental Design with Marketing Research Applications," *Journal of Marketing Research*, (November), 545-557.

Leigh, Thomas W., David B. MacKay, and John O. Summers (1984), "Reliability and Validity of Conjoint Analysis and Self-Explicated Weights: A Comparison," *Journal of Marketing Research*, pp. 456-462.

Lenk, Peter J., Wayne S. DeSarbo, Paul E. Green, and Martin R. Young (1996), "Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs," *Marketing Science*, 15, 2, 173-191.

Liechty, John, Venkatram Ramaswamy, Steven Cohen (2001), "Choice-Menus for Mass Customization: An Experimental Approach for Analyzing Customer Demand With an Application to a Web-based Information Service," *Journal of Marketing Research*, 38, 2, (May).

Louviere, Jordan J., David A. Hensher, and Joffre D. Swait (2000), *Stated Choice Methods: Analysis and Application*, (New York, NY: Cambridge University Press), pp. 354-381.

Mahajan, Vijay and Jerry Wind (1992), "New Product Models: Practice, Shortcomings and Desired Improvements," *Journal of Product Innovation Management*, 9, 128-139.

______ (1986), “An Approach to the Measurement of Consumer Preferences Using Limited Information,” *Journal of Marketing Research*, 23, (February), 33-40.

McGrath, Michael E. (1996), *Setting the Pace in Product Development: A guide to Product and Cycle-Time Excellence*, (Boston, MA: Butterworth-Heinemann).

Montgomery, David B. and Dick R. Wittink (1980), “The Predictive Validity of Conjoint Analysis for Alternative Aggregation Schemes,” in *Market Measurement and Analysis: Proceedings of the 1979 ORSA/TIMS Conference on Marketing*, David B. Montgomery and Dick R. Wittink, eds, Cambridge, MA: Marketing Science Institute, 298-309.

Moore, William L. (1980), “Levels of Aggregation in Conjoint Analysis: An Empirical Comparison,” *Journal of Marketing Research*, pp. 516-523.

______ and Richard J. Semenik (1988), “Measuring Preferences with Hybrid Conjoint Analysis: The Impact of a Different Number of Attributes in the Master Design,” *Journal of Business Research*, pp. 261-274.

Nadilo, Rudy (1999), “On-line Research: The Methodology for the Next Millennium,” *Advertising Research Foundation Journal*, (Spring). Available at www.greenfield.com.

Nesterov, Y. and A. Nemirovskii (1994), “Interior-Point Polynomial Algorithms in Convex Programming,” SIAM, Philadelphia.

Nowlis, Stephen M. and Itamar Simonson (1997), “Attribute-Task Compatibility as a Determinant of Consumer Preference Reversals,” *Journal of Marketing Research*, 36, (May), 205-218.

Orme, Bryan (1999), “ACA, CBC, of Both?: Effective Strategies for Conjoint Research,” Working Paper, Sawtooth Software, Sequim, WA.

______, M. I. Alpert, and E. Christensen (1997), “Assessing the Validity of Conjoint Analysis – Continued,” *Sawtooth Software Conference Proceedings*, 209-225. <Need>

Page, Albert L. and Harold F. Rosenbaum (1987), “Redesigning Product Lines with Conjoint Analysis: How Sunbeam Does It,” *Journal of Product Innovation Management*, 4, 120-137.

Park, C. Whan, Sung Youl Jun, and Deborah Macinnis (2000), “Choosing What I Want versus Rejecting What I Do Not Want: An Application of Decision Framing to Product Option Choice Decisions,” *Journal of Marketing Research*, 37, 2, (May), 187-202.

Parker B. R. and V. Srinivasan (1976), “A Consumer Preference Approach to the Planning of Rural Primary Health-Care Facilities,” *Operations Research*, 24, 5, (September-October), 991-1025.

Robinson, P. J. (1980), “Applications of Conjoint Analysis to Pricing Problems,” in *Market Measurement and Analysis: Proceedings of the 1979 ORSA/TIMS Conference on Marketing*, David B. Montgomery and Dick Wittink, eds., (Cambridge, MA: Marketing Science Institute), 183-205. <Need>

Sawtooth Software, Inc. (1999), “CVA: Full-Profile Conjoint System from Sawtooth Software,” *CVA Manual*, (Sequim, WA: Sawtooth Software, Inc.)

______ (1996), “ACA System: Adaptive Conjoint Analysis,” *ACA Manual*, (Sequim, WA: Sawtooth Software, Inc.)

______ (1999), "The ACA/HB Module for Hierarchical Bayes Estimation," *ACA/HB Manual*, (Sequim, WA: Sawtooth Software, Inc.)

Sawyer, Alan G. (1975), "Demand Artifacts in Laboratory Experiments in Consumer Research," *Journal of Consumer Research*, 1. (March), 20-30.

Sheth, Jagdish N. and Wayne W. Talarzyk (1972), "Perceived Instrumentality and Value Importance as Determinants of Attitudes," *Journal of Marketing Research*, 9, 6-9.

Shimp, Terry, Eva Hyatt, and David Snyder (1991), "A critical Appraisal of Demand Artifacts in Laboratory Experiments in Consumer Research," *Journal of Consumer Research,* 18, (December), 273-283.

Simmons, Carolyn J., Barbara A. Bickart, and John G. Lynch, Jr. (1993), "Capturing and Creating Public Opinion in Survey Research," *Journal of Consumer Research*, 30, (September), 316-329.

Smith, Preston G. and Donald G. Reinertsen (1998), *Developing Products in Half the Time, 2E*, (New York, NY: John Wiley & Sons, Inc.).

Sonnevend, G. (1985a), "An 'Analytic' Center for Polyhedrons and New Classes of Global Algorithms for Linear (Smooth, Convex) Programming," *Proceedings of the 12th IFIP Conference on System Modeling and Optimization*, Budapest.

______ (1985b), "A New Method for Solving a Set of Linear (Convex) Inequalities and its Applications for Identification and Optimization," Preprint, Department of Numerical Analysis, Institute of Mathematics, Eőtvős University, Budapest, 1985.

Srinivasan, V. (1988), "A Conjunctive-Compensatory Approach to The Self-Explication of Multiattributed Preferences," *Decision Sciences*, pp. 295-305.

______ and Chan Su Park (1997), "Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement," *Journal of Marketing Research*, 34, (May), 286-291.

______ and Allen D. Shocker(1973), "Estimating the Weights for Multiple Attributes in a Composite Criterion Using Pairwise Judgments," *Psychometrika*, 38, 4, (December), 473-493.

______ and ______ (1973), "Linear Programming Techniques for Multidimensional Analysis of Preferences," *Psychometrika*, 38, 3, (September), 337-369.

Tessler, Amy, Norm Wada, and Robert L. Klein (1993), "QFD at PG&E," *Transactions from The Fifth Symposium on Quality Function Deployment,* (June).

Toubia, Olivier, Duncan Simester, John R. Hauser (2001), "Fast Polyhedral Adaptive Conjoint Estimation," working paper, Cambridge, MA: Center for Innovation in Product Development, MIT, (May).

Tourangeau, Roger, Lance J. Rips, and Kenneth A. Rasinski (2000), *The Psychology of Survey Response*, (New York, NY: Cambridge University Press).

Tybout, Alice M. and John R. Hauser (1981), "A Marketing Audit Using a Conceptual Model of Consumer Behavior: Application and Evaluation," *Journal of Marketing*, Vol. 45, No. 3, (Summer), 81-101.

Ulrich, Karl T. and Steven D. Eppinger (2000), *Product Design and Development*, (New York, NY: McGraw-Hill, Inc.).

Urban, Glen L. and Gerald M. Katz, "Pre-Test Market Models: Validation and Managerial Implications," *Journal of Marketing Research*, Vol. 20 (August 1983), 221-34.

Vaidja, P. (1989), "A Locally Well-Behaved Potential Function and a Simple Newton-Type Method for Finding the Center of a Polytope," in: N. Megiddo, ed., *Progress in Mathematical Programming: Interior Points and Related Methods*, Springer: New York, 79-90.

Wilkie, William L. and Edgar A. Pessemier (1973), "Issues in Marketing's Use of Multi-attribute Attitude Models," *Journal of Marketing Research*, 10, (November), 428-441.

Willkie, Joseph., Adams, Christine and Azra Girnius (1999), "First Systematic Comparison Between Mall-Intercept and Internet Interviewing," European Society for Opinion and Marketing Research, Worldwide Internet Conference, London UK, (February 21).

Wind, Jerry, Paul E. Green, Douglas Shifflet, and Marsha Scarbrough (1989), "Courtyard by Marriott: Designing a Hotel Facility with Consumer-Based Marketing Models," *Interfaces*, pp. 25-47.

Wittink, Dick R. and Trond Bergestuen (2001), "Forecasting with Conjoint Analysis," *Principles of Forecasting: A Handbook for Researchers and Practitioners*, J. Scott Armstrong, ed., (Norwell, MA: Kluwer Academic Publishers),

______ and Philippe Cattin (1989), "Commercial Use of Conjoint Analysis: An Update," *Journal of Marketing*, pp. 91-96.

______ and David B. Montgomery (1979), "Predictive Validity of Trade-off Analysis for Alternative Segmentation Schemes," *1979 AMA Educators' Conference Proceedings,* Neil Beckwith ed., Chicago, IL: American Marketing Association.

Wright, Peter and Mary Ann Kriewall (1980), "State-of-Mind Effects on Accuracy with which Utility Functions Predict Marketplace Utility," *Journal of Marketing Research*, 17, (August), 277-293.