Fast Polyhedral Adaptive Conjoint Estimation

Olivier Toubia • Duncan I. Simester • John R. Hauser • Ely Dahan
Sloan School of Management, Massachusetts Institute of Technology, E56-305, 38 Memorial Drive, Cambridge, Massachusetts 02142
Anderson School, University of California at Los Angeles, 110 Westwood Plaza, B-514, Los Angeles, California 90095
toubia@mit.edu. • simester@mit.edu. • jhauser@mit.edu. • edahan@ucla.edu

We propose and test new adaptive question design and estimation algorithms for partial-profile conjoint analysis. Polyhedral question design focuses questions to reduce a feasible set of parameters as rapidly as possible. Analytic center estimation uses a centrality criterion based on consistency with respondents' answers. Both algorithms run with no noticeable delay between questions.

We evaluate the proposed methods relative to established benchmarks for question design (random selection, D-efficient designs, adaptive conjoint analysis) and estimation (hierarchical Bayes). Monte Carlo simulations vary respondent heterogeneity and response errors. For low numbers of questions, polyhedral question design does best (or is tied for best) for all tested domains. For high numbers of questions, efficient fixed designs do better in some domains. Analytic center estimation shows promise for high heterogeneity and for low response errors; hierarchical Bayes for low heterogeneity and high response errors. Other simulations evaluate hybrid methods, which include self-explicated data.

A field test (330 respondents) compared methods on both internal validity (holdout tasks) and external validity (actual choice of a laptop bag worth approximately $100). The field test is consistent with the simulation results and offers strong support for polyhedral question design. In addition, marketplace sales were consistent with conjoint-analysis predictions.
(*New Product Research; Measurement; Internet Marketing; Estimation and Other Statistical Techniques*)

1. Polyhedral Methods for Conjoint Analysis

We propose and test (1) a new adaptive question-design method that attempts to reduce respondent burden while simultaneously improving accuracy and (2) a new estimation procedure based on centrality concepts. For each respondent the question-design method dynamically adapts the design of the next question using that respondent's answers to previous questions. Because the methods make full use of high-speed computations and adaptive, customized local Web pages, they are ideally suited for Web-based panels. The adaptive method interprets question design as a mathematical program and estimates the solution to the program using recent developments based on the interior points of polyhedra. The estimation method also relies on interior point techniques and is designed to provide robust estimates from relatively few questions. The question design

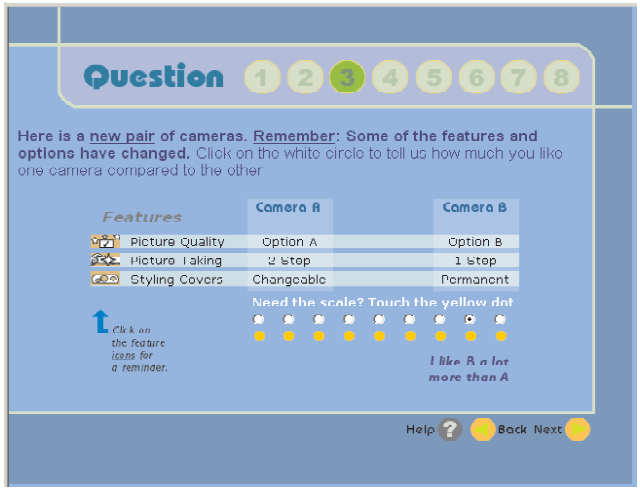
0732-2399/03/2203/0273
1526-548X electronic ISSN

and estimation methods are modular and can be evaluated separately and/or combined with a range of existing methods.

Adapting question design within a respondent, using that respondent's answers to previous questions, is a difficult dynamic optimization problem. Adaptation *within* respondents should be distinguished from techniques that adapt *across* respondents. Sawtooth Software's adaptive conjoint analysis (ACA) is the only published method of which we are aware that attempts to solve this problem (Johnson 1987, 1991). In contrast, aggregate customization methods, such as the Huber and Zwerina (1996), Arora and Huber (2001), and Sandor and Wedel (2001, 2002) algorithms, adapt designs across respondents based on either pretests or Bayesian priors.

ACA uses a data-collection format known as metric paired-comparison questions and relies on balancing utility between the pairs subject to orthogonality and feature balance. We provide an example of a metric-paired comparison question in Figure 1. To date, aggregate customization methods have focused on a stated-choice data-collection format known as choice-based conjoint (CBC; e.g. Louviere et al. 2000). Polyhedral methods can be used to design either metric-paired-comparison questions or choice-based questions. In this paper we focus on metric-paired-comparison questions because this is one of the most widely used and applied data-collection formats for conjoint analysis (Green et al. 2001, p. S66; Ter Hofstede et al. 2002, p. 259). In addition, metric paired-comparison questions are common in computer-aided interviewing, have proven reliable in previous studies (Reibstein et al. 1988, Urban and Katz 1983), provide interval-scaled data with strong transitivity properties (Hauser and Shugan 1980), provide valid and reliable parameter estimates (Leigh et al. 1984), and enjoy wide use in practice and in the literature (Wittink and Cattin 1989). We are extending polyhedral methods to CBC formats (Toubia et al. 2003).

Figure 1 **Metric Paired-Comparison Format for I-Zone Camera Redesign**



Our goal is an initial evaluation of polyhedral methods relative to existing methods under a variety of empirically relevant conditions. We do not expect that any one method will always outperform the benchmarks, nor do we intend that our findings be interpreted as criticism of any of the benchmarks. Our findings indicate that polyhedral methods have the potential to enhance the effectiveness of existing conjoint methods by providing new capabilities that complement existing methods.

Because the methods are new and adopt a different estimation philosophy, we use Monte Carlo experiments to explore their properties. The Monte Carlo experiments explore the conditions under which polyhedral methods are likely to do better or worse than extant methods. We demonstrate practical domains where polyhedral methods show promise relative to a representative set of widely applied and studied methods. The findings also highlight opportunities for future research by illustrating domains where improvements are necessary and/or where extant methods are likely to remain superior.

We also undertake a large-scale empirical test involving a real product—a laptop computer bag worth approximately $100. Respondents first completed a series of Web-based conjoint questions chosen by one of three question-design methods (the methods were assigned randomly). After a filler task, respondents in the study were given $100 to spend on a choice set of five bags. Respondents received their chosen bag together with the difference in cash between the price of their chosen bag and the $100. We compare question-design and estimation 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 which bag respondents later chose to purchase using their $100.

The paper is structured as follows. We begin by describing polyhedral question design and analytic center estimation for metric paired-comparison tasks. Detailed mathematics are provided in Appendix 1 and open-source code is available from http://mktsci.pubs.informs.org. We next describe the design and results of the Monte Carlo experiments. Finally, we describe the field test and the comparative results. We close with a description of the launch of the laptop bag, a summary of the findings, and suggestions for future research.

2. Polyhedral Question Design and Estimation

We begin with a conceptual description that highlights the geometry of the conjoint-analysis parameter space. We illustrate the concepts with a three-parameter problem because three-dimensional spaces are easy to visualize and explain. The methods generalize easily to realistic problems that contain 10, 20, or even 100 product features. Indeed, relative to existing methods, the polyhedral methods are proposed for larger numbers of product features. By a parameter, we refer to a partworth that needs to be estimated. For example, 20 features with 2 levels each require 20 parameters because we can scale to zero the partworth of the least preferred feature. Similarly, 10 three-level features also require 20 parameters. Interactions among features require still more parameters.

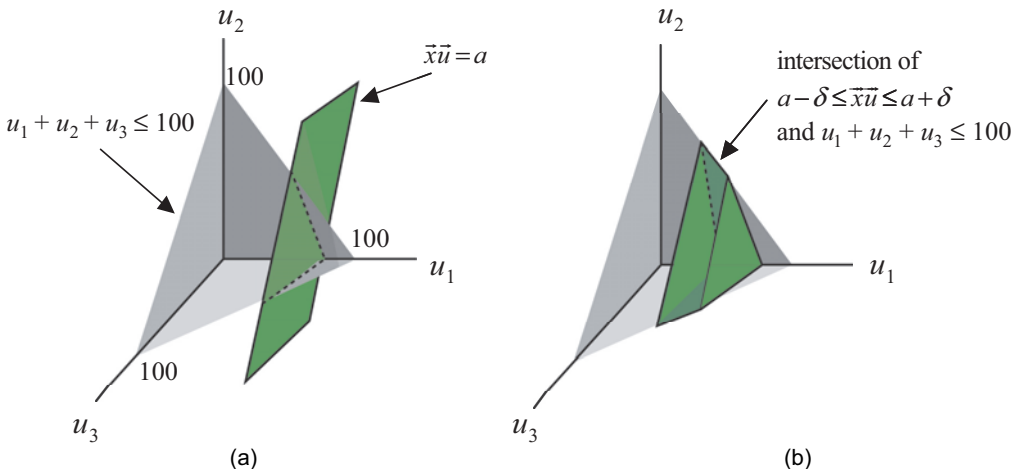
Suppose that we have three features of an instant camera: picture quality, picture taking (two-step versus one-step), and styling covers (changeable versus permanent). If we scale the least desirable level of each feature to zero, we have three nonnegative parameters to estimate—u_1, u_2, and u_3—reflecting the additional utility (partworth) associated with the most desirable level of each feature.[1] The measurement scale on which the questions are asked imposes natural boundary conditions. For example, the sum of the partworths of Camera A minus the sum of the partworths of Camera B can be at most equal to the maximum scale difference. In practice, the partworths have only relative meaning, so scaling allows us to impose a wide range of boundary conditions without loss of generality. Therefore, to better visualize the algorithm, we impose a constraint that the sum of the parameters does not exceed some large number (e.g., 100). Under this constraint, prior to any data collection, the feasible region for the parameters is the three-dimensional bounded polyhedron in Figure 2a.

Suppose that we ask the respondent to evaluate a pair of profiles that vary on one or more features and the respondent says (1) that he or she prefers profile C_1 to profile C_2, and (2) provides a rating, a, to indicate the strength of his or her preference. Assuming for the moment that the respondent answers without error, this introduces an equality constraint that the utility associated with profile C_1 exceeds the utility of C_2 by an amount equal to the rating. If we define $\vec{u} = (u_1, u_2, u_3)^T$ as the 3×1 vector of parameters, $\vec{z}_l$ as the 1×3 vector of product features for the left profile, and $\vec{z}_r$ as the 1×3 vector of product features for the right profile, then, for additive utility, this equality constraint can be written as $\vec{z}_l\vec{u} - \vec{z}_r\vec{u} = a$. We can use geometry to characterize what we have learned from this response.

Specifically, we define $\vec{x} = \vec{z}_l - \vec{z}_r$ such that $\vec{x}$ is a 1×3 vector describing the difference between the two profiles in the question. Then, $\vec{x}\vec{u} = a$ defines a hyperplane through the polyhedron in Figure 2a. The only feasible values of $\vec{u}$ are those that are in the intersection of this hyperplane and the polyhedron. The new feasible set is also a polyhedron, but it is reduced by one dimension (two dimensions rather than three dimensions). Because smaller polyhedra mean fewer parameter values are feasible, questions that reduce

[1] In this example, we assume preferential independence which implies an additive utility function. We can handle interactions by relabeling features. For example, a 2×2 interaction between two features is equivalent to one four-level feature. We hold to this convention throughout the paper.

Figure 2 Respondent's Answers Affect the Feasible Region



Notes. (a) Metric rating without error. (b) Metric rating with error.

the size of the initial polyhedron as fast as possible lead to more precise estimates of the parameters.

However, in any real problem we expect a respondent's answer to contain error. We can model this error as a probability density function over the parameter space (as in standard statistical inference). Alternatively, we can incorporate imprecision in a response by treating the equality constraint $\vec{x}\vec{u} = a$ as a set of two inequality constraints: $a - \delta \leq \vec{x}\vec{u} \leq a + \delta$. In this case, the hyperplane defined by the question-answer pair has "width." The intersection of the initial polyhedron and the "fat" hyperplane is now a three-dimensional polyhedron as illustrated in Figure 2b.

When we ask more questions we constrain the parameter space further. Each question, if asked carefully, will result in a hyperplane that intersects a polyhedron resulting in a smaller polyhedron—a "thin" region in Figure 2a or a "fat" region in Figure 2b. Each new question-answer pair slices the polyhedron in Figure 2a or 2b yielding more precise estimates of the parameter vector $\vec{u}$.

We incorporate prior information about the parameters by imposing constraints on the parameter space. For example, if u_m and u_h are the medium and high levels, respectively, of a feature, then we impose the constraint $u_m \leq u_h$ on the polyhedron. Previous research suggests that these types of constraints enhance estimation (Johnson 1999, Srinivasan and Shocker 1973). We now examine question design for metric paired-comparison data by dealing first with the case in which subjects respond without error (Figure 2a). We then describe how to modify the algorithm to handle error (e.g., Figure 2b).

Selecting Questions to Shrink the Feasible Set Rapidly

The question-design task describes the design of the profiles that respondents are asked to compare. Questions are more informative if the answers allow us to estimate partworths more quickly. For this reason, we select the respondent's next question in a manner that is likely to reduce the size of the feasible set (for that respondent) as fast as possible.

Consider for a moment a 20-dimensional problem (without errors in the answers). As in Figure 2a, a question-based constraint reduces the dimensionality by one. That is, the first question reduces a 20-dimensional set to a 19-dimensional set; the next question reduces this set to an 18-dimensional set, and so on. After the twelfth question, for example, we reach an eight-dimensional set: 8 dimensions = 20 parameters − 12 questions. Without further restriction, the feasible parameters are generally not unique—any point in the eight-dimensional set (polyhedron) is still feasible. However, the eight-dimensional set might be quite small, and we might have a very good idea of the partworths. For example, the first 12 questions

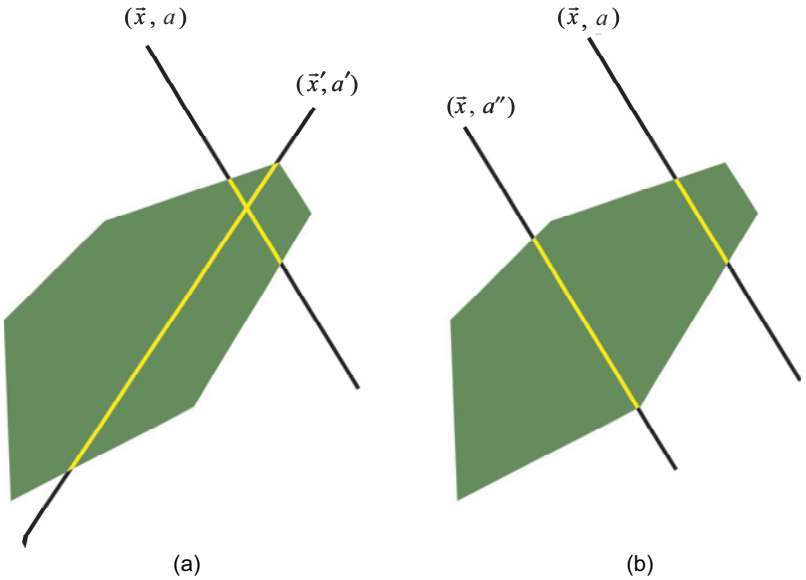
might be enough to tell us that some features—say picture quality, styling covers, and battery life—have large partworths, while other features—say folding capability, light selection, and film ejection method—have very small partworths. If this holds across respondents, then during an early phase of a product development process, the product development team might feel they have enough information to focus on the key features.

Although the polyhedral algorithm is designed for high-dimensional spaces, it is hard to visualize 20-dimensional polyhedra. Instead, we illustrate the polyhedral question-design method in a situation where the remaining feasible set is easy to visualize. Specifically, by generalizing our notation slightly to q questions and p parameters, we define $\vec{a}$ as the $q \times 1$ vector of answers and X as the $q \times p$ matrix with rows equal to $\vec{x}$ for each question (recall that $\vec{x}$ is a $1 \times p$ vector). Then the respondent's answers to the first q questions define a $(p-q)$-dimensional hyperplane given by the equation $X\vec{u} = \vec{a}$. This hyperplane intersects the initial p-dimensional polyhedron to give us a $(p-q)$-dimensional polyhedron. In the example of $p = 20$ parameters and $q = 18$ questions, the result is a two-dimensional polyhedron that is easy to visualize. One such two-dimensional polyhedron is illustrated in Figure 3.

Our task is to select questions that reduce the two-dimensional polyhedron as fast as possible. Mathematically, we select a new question vector, $\vec{x}$, and the respondent answers this question with a new rating, a. We add the new question vector as the last row of the question matrix and we add the new answer as the last row of the answer vector. While everything is really happening in p-dimensional space, the net result is that the new hyperplane will intersect the two-dimensional polyhedron in a line segment (i.e., a one-dimensional polyhedron). The slope of the line will be determined by $\vec{x}$ and the intercept by a. We illustrate two potential question-answer pairs in Figure 3a. The slope of the line is determined by the question, the specific line by the answer, and the remaining feasible set by the line segment within the polyhedron. In Figure 3a one of the question-answer pairs $(\vec{x}, a)$ reduces the feasible set more rapidly than the other question-answer pair $(\vec{x}', a')$. Figure 3b repeats a question-answer pair $(\vec{x}, a)$ and illustrates an alternative answer to the same question $(\vec{x}, a'')$.

If the polyhedron is elongated, as in Figure 3, then in most cases, questions that imply line segments

Figure 3 Choice of Question (2-Dimensional Slice)



Notes. (a) Two question-answer pairs. (b) One question, two potential answers.

perpendicular to the longest "axis" of the polyhedron are questions that result in the smallest remaining feasible sets. Also, because the longest "axis" is in some sense a bigger target, it is more likely that the respondent's answer will select a hyperplane that intersects the polyhedron. From analytic geometry we know that hyperplanes (line segments in Figure 3) are perpendicular to their defining vectors ($\vec{x}$). Thus, we can reduce the feasible set as fast as possible (and make it more likely that answers are feasible) if we choose question vectors that are parallel to the longest "axis." For example, both line segments based on $\vec{x}$ in Figure 3b are shorter than the line segment based on $\vec{x}'$ in Figure 3a.

If we can develop an algorithm that works in any p-dimensional space, then we can generalize this intuition to any question, q, such that $q \leq p$. After receiving answers to the first q questions, we could find the longest vector of the (p-q)-dimensional polyhedron of feasible parameter values. We could then ask the question based on a vector that is parallel to this "axis." The respondent's answer creates a hyperplane that intersects the polyhedron to produce a new polyhedron. We address later the cases where respondents' answers contain error and where $q > p$.

Centrality Estimation

Polyhedral geometry also gives us a means to estimate the parameter vector, $\vec{u}$, when $q \leq p$. Recall that, after question q, any point in the remaining polyhedron is consistent with the answers the respondent has provided. If we impose a diffuse prior that any feasible point is equally likely, then we would like to select the point that minimizes the expected error. This point is the center of the feasible polyhedron, or more precisely, the polyhedron's center of gravity. The smaller the feasible set, either due to better question design or more questions (higher q), the more precise the estimate. If there were no respondent errors, then the estimate would converge to its true value when $q = p$ (the feasible set becomes a single point, with zero dimensionality). For $q > p$ the same point would remain feasible. As we discuss below, this changes when responses contain error.

This technique of estimating partworths from the center of a feasible polyhedron is related to that proposed by Srinivasan and Shocker (1973, p. 350) who suggest using a linear program to find the "innermost" point that maximizes the minimum distance from the hyperplanes that bound the feasible set. Philosophically, the proposed polyhedral method makes use of the information in the constraints and then takes a central estimate based on what is still feasible. Carefully chosen questions shrink the feasible set rapidly. We then use a centrality estimate that has proven to be a surprisingly good approximation in a variety of engineering problems. More generally, the centrality estimate is similar in some respects to the proven robustness of linear models, and in some cases, to the robustness of equally weighted models (Dawes and Corrigan 1974, Einhorn 1971, Huber 1975, Moore and Semenik 1988, Srinivasan and Park 1997).

Interior-Point Algorithms and the Analytic Center of a Polyhedron

To select questions and obtain intermediate estimates, the proposed heuristics require that we solve two nontrivial mathematical programs. First, we must find the longest "axis" of a polyhedron (to select the next question) and second, we must find the polyhedron's center of gravity (to provide a centrality estimate). If we were to define the longest "axis" of a polyhedron as the longest line segment in the polyhedron, then one method to find the longest "axis" would be to enumerate the vertices of the polyhedron and compute the distances between the vertices. However, solving this problem requires checking every extreme point, which is computationally intractable (Gritzmann and Klee 1993). In practice, solving the problem would impose noticeable delays between questions. Also, the longest line segment in a polyhedron may not capture the concept of a longest "axis." Finding the center of gravity of the polyhedron is even more difficult and computationally demanding.

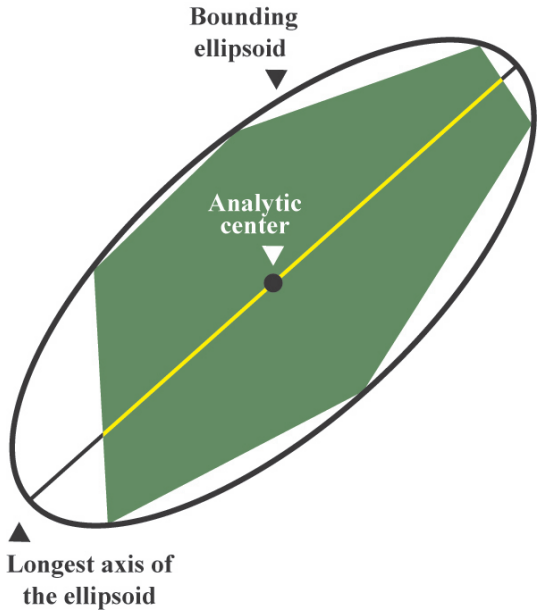
Fortunately, recent work in the mathematical programming literature has led to extremely fast algorithms based on projections within the interior of polyhedrons (much of this work started with Karmarkar 1984). Interior-point algorithms are now used routinely to solve large problems and have

spawned many theoretical and applied generalizations. One such generalization uses bounding ellipsoids. In 1985, Sonnevend demonstrated that the shape of a bounded polyhedron can be approximated by proportional ellipsoids, centered at the "analytic center" of the polyhedron. The analytic center is the point in the polyhedron that maximizes the geometric mean of the distances to the boundaries of the polyhedron. It is a central point that approximates the center of gravity of the polyhedron, and finds practical use in engineering and optimization. Furthermore, the axes of the ellipsoids are well-defined and intuitively capture the concept of an "axis" of a polyhedron. For more details see Freund (1993), Nesterov and Nemirovskii (1994), Sonnevend (1985a, b), and Vaidja (1989).

Polyhedral Question Design and Analytic Center Estimation

We illustrate the proposed process in Figure 4, using the same two-dimensional polyhedron depicted in Figure 3. The algorithm proceeds in four steps. We first find a point in the interior of the polyhedron. This is a simple linear programming (LP) problem and runs quickly. Then, following Freund (1993) we use Newton's method to make the point more central. This is a well-formed problem and converges quickly to yield the analytic center as illustrated by the black dot in Figure 4. We next find a bounding ellipsoid based on a formula that depends on the analytic center and the question-matrix, X. We then find the longest axis of the ellipsoid (diagonal line in Figure 4) with a quadratic program that has a closed-form solution. The next question, $\vec{x}$, is based on the vector most nearly parallel to this axis. A formal (mathematical) description of each step is provided in Appendix 1.

Figure 4 Bounding Ellipsoid and the Analytic Center (2-Dimensions)



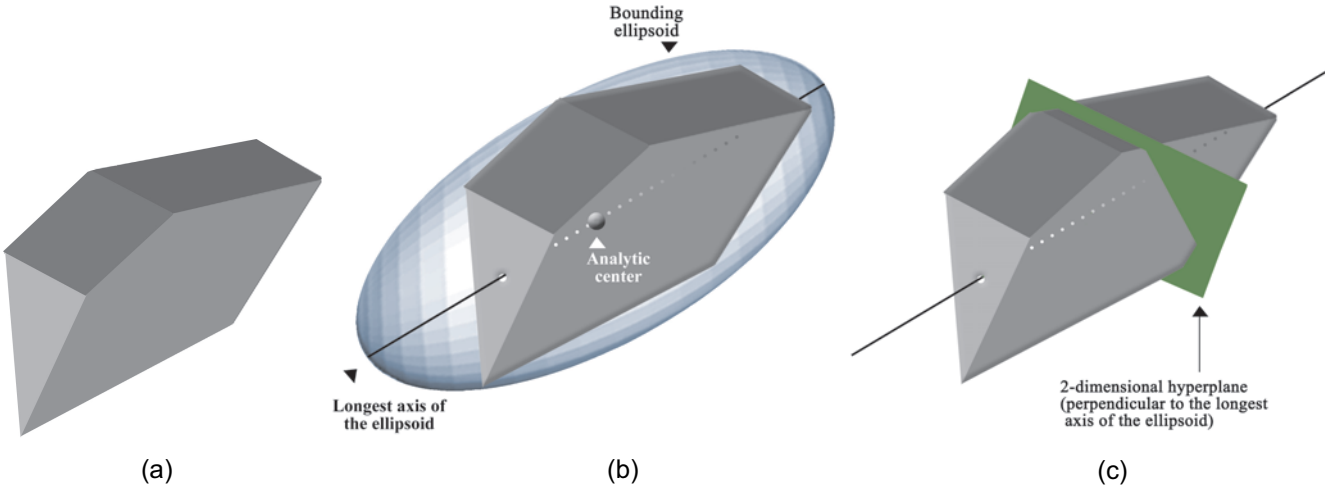
Analytically, this algorithm works well in higher dimensional spaces. For example, Figure 5 illustrates the algorithm when $(p-q)=3$, where we reduce a three-dimensional feasible set to a two-dimensional feasible set. Figure 5a illustrates a polyhedron based on the first q questions. Figure 5b illustrates a bounding three-dimensional ellipsoid, the longest axis of that ellipsoid, and the analytic center. The longest axis defines the question that is asked next, which in turn defines the slope of the hyperplane that intersects the polyhedron. One such hyperplane is shown in Figure 5c. The respondent's answer locates the specific hyperplane. The intersection of the selected hyperplane and the three-dimensional polyhedron is a new two-dimensional polyhedron, such as that in Figure 4. This process applies (in higher dimensions) from the first question to the pth question. For example, the first question implies a hyperplane that cuts the first p-dimensional polyhedron such that the intersection yields a $(p-1)$-dimensional polyhedron.

The polyhedral algorithm runs extremely fast. We have implemented the algorithm for the Web-based empirical test described later in this paper. Based on this example, with 10 two-level features, respondents noticed no delay in question design nor any difference in speed versus a fixed design. For a demonstration see the Website referenced in the Acknowledgments.

Inconsistent Responses and Error-Modeling

Figures 2 through 5 illustrate the geometry when respondents answer without error. However, real respondents are unlikely to be perfectly consistent. It is more likely that, for some $q < p$, the respondent's answers will be inconsistent and the polyhedron will become empty. That is, we will no longer be able to

Figure 5 Question Design with a 3-Dimensional Polyhedron



Notes. (a) Polyhedron in 3 dimensions. (b) Bounding ellipsoid, analytic center, and longest axis. (c) Example hyperplane determined by question vector and respondent's answer.

find any parameters, $\vec{u}$, that satisfy the equations that define the polyhedron, $X\vec{u} = \vec{a}$. Thus, for real applications, we extend the polyhedral algorithm to address response errors. Specifically, we adjust the polyhedron in a minimal way to ensure that some parameter values are still feasible. We do this by modeling errors, $\vec{\delta}$, in the respondent's answers such that $\vec{a} - \vec{\delta} \leq X\vec{u} \leq \vec{a} + \vec{\delta}$ (recall Figure 2b). We then choose the minimum errors such that these constraints are satisfied. This same modification covers estimation for the case of $q > p$. Appendix 1 provides the mathematical program (OPT4) that we use to estimate $\vec{u}$ and $\vec{\delta}$. The algorithm is easily modified to incorporate alternative error formulations, such as least-squares or minimum sum of absolute deviations, rather than this "minimax" criterion.[2] Exploratory simulations suggest that the algorithm is robust to the choice of error criterion.

To implement this policy for analytic center estimation, we use a two-stage algorithm. In the first stage we treat the responses as if they occurred without error—the feasible polyhedron shrinks rapidly and the analytic center is a working estimate of the true parameters. However, as soon as the feasible set becomes empty, we adjust the constraints by adding or subtracting "errors," where we choose the minimum errors, $\|\vec{\delta}\|$, for which the feasible set is nonempty. The analytic center of the new polyhedron becomes the working estimate and $\vec{\delta}$ becomes an index of response error.[3] As with all our heuristics, the accuracy of our error-modeling method is tested with simulation. While estimates based on this heuristic seem to converge for the domains that we test (reduce mean errors, no measured bias), we recognize the need for further exploration of this heuristic, together with the development of a formal error theory.

Addressing Other Practical Implementation Issues

Implementation raises several additional issues. Alternative solutions to these issues may yield more or less accurate parameter estimates, and so the performance of the polyhedral methods in the validation tasks are lower bounds on the performance of this class of methods.

[2] Technically, the minimax criterion is called the "∞-norm." To handle least-squares errors we use the "2-norm" and to handle average absolute errors we use the "1-norm." Either is a simple modification to OPT4 in Appendix 1.

[3] By construction, $\vec{\delta}$ grows (weakly) with the number of questions, q, thus a better measure of fit might be mean error, the error divided by the number of questions—an analogy to mean-squared error in regression.

Product Profiles with Discrete Features. In most conjoint analysis problems the features are specified at discrete levels, as in Figure 1. This constrains the elements of the $\vec{x}$ vector to be $1, -1, 0$, or 0, depending on whether the left profile, the right profile, neither profile, or both profiles have the "high" feature, respectively. In this case we choose the vector that is most nearly parallel to the longest axis of the ellipsoid. Because we can always recode multilevel features or interacting features as binary features, the geometric insights still hold even if we otherwise simplify the algorithm.

Restrictions on Question Design. For a p-dimensional problem we may wish to vary fewer than p features in any paired-comparison question. For example, Sawtooth Software (1996, p. 7) suggests that: "Most respondents can handle three attributes after they've become familiar with the task. Experience tells us that there does not seem to be much benefit from using more than three attributes." We incorporate this constraint by restricting the set of questions over which we search when finding a question-vector that is parallel to the longest axis of the ellipse.

First Question. Unless we have prior information before any question is asked, the initial polyhedron of feasible utilities is defined by the boundary constraints. If the boundary constraints are symmetric, the polyhedron is also symmetric and the polyhedral method offers little guidance for the choice of the first question. In these situations we choose the first question for each respondent so that it helps improve estimates of the population means by balancing how often each feature level appears in the set of questions answered by all respondents. In particular, for the first question presented to each respondent we choose feature levels that appeared infrequently in the questions answered by previous respondents.

Question Design When the Parameter Set Becomes Infeasible. Analytic center estimation is well-defined when the parameter set becomes infeasible, but question design is not. Thus, in the simulations we use a random question design heuristic when the parameter set is infeasible.[4] This provides a lower bound on what might be achieved.

Programming. The optimization algorithms used for the simulations are written in Matlab and are available at http://mitsloan.mit.edu/vc. This Web site contains (1) open source code to implement the methods described in this paper, (2) open source code for the simulations described in this paper, (3) detailed instructions for implementing the method, (4) worked examples, (5) demonstrations of Web-based questionnaires, (6) raw empirical data, and (7) related papers on Web-based interviewing methods. The open source code, detailed instructions data and worked examples are also available on the *Marketing Science* Web site.

3. Monte Carlo Simulations

Polyhedral methods for conjoint analysis are new and untested. Although interior-point algorithms and the centrality criterion have been successful in many engineering problems, we are unaware of any prior application to marketing problems. Thus, we turn first to Monte Carlo experiments to identify circumstances in which polyhedral methods may contribute to the effectiveness of current methods. Monte Carlo simulations offer at least three advantages for the *initial* test of a new method. First, they facilitate comparison of different techniques in a range of domains such as varying levels of respondent heterogeneity and response accuracy. We can also evaluate combinations of the techniques, for example, mixing polyhedral question design with extant estimation methods. Second, simulations resolve the issue of identifying the correct answer. In studies involving actual customers, the true partial utilities are unobserved. In simulations the true partial utilities are constructed so that we can compare how well alternative methods identify the true utilities from noisy

[4] Earlier implementations, including the field test, used ACA question design when the parameter set became infeasible. Further analysis revealed that it is better to switch to random question design than ACA question design when the parameter set becomes infeasible. This makes the performance of polyhedral question design in the field test conservative.

responses. Finally, other researchers can readily replicate the findings. However, simulations do not guarantee that real respondents behave as simulated nor do they reveal which domain is likely to best summarize field experience. Thus, following the simulations, we examine a field test that matches one of the simulated domains.

Many papers have used the relative strengths of Monte Carlo experiments to study conjoint techniques, providing insights on interactions, robustness, continuity, feature correlation, segmentation, new estimation methods, new data-collection methods, post-analysis with hierarchical Bayes methods, and comparisons of ACA, CBC, and other conjoint methods. Although we focus on specific benchmarks, there are many comparisons in the literature of these benchmarks to other methods (see reviews and citations in Green 1984; Green et al. 2001, 2002; Green and Srinivasan 1978, 1990; Hauser and Rao 2003; Moore 2003).

We test polyhedral question design versus three question design benchmarks and analytic center estimation versus hierarchical Bayes estimation. The initial simulations vary respondent heterogeneity, accuracy of respondent answers, and the number of questions. In a second set of simulations we also consider the role of self-explicated responses and vary the accuracy of self-explicated responses.

Respondent Heterogeneity, Response Errors, and Number of Questions

We focus on a design problem involving 10 features, where a product development team is interested in learning the incremental utility contributed by each feature. We follow convention and scale to zero the partworth of the low level of a feature and, without loss of generality, bound it by 100. This results in a total of 10 parameters to estimate ($p = 10$). We anticipate that the polyhedral methods are particularly well-suited to solving problems in which there are a large number of parameters relative to the number of responses from each individual ($q < p$). Thus, we vary the number of questions from slightly less than the number of parameters ($q = 8$) to comfortably more than the number of parameters ($q = 16$).

We simulate each respondent's partworths by drawing independently and randomly from a normal distribution with mean 50 and variance σ_u^2, truncated to the range. We explored the sensitivity of the findings to this specification by testing different methods of drawing partworths, including beta distributions that tend to yield more similar partworths (inverted-U shape distributions), more diverse partworths (U-shaped distributions), or moderately diverse partworths (uniform distributions). Sensitivity analyses for key findings did not suggest much variation. Nonetheless, this is an important area for more systematic future research. By manipulating the standard deviation of the normal distribution we explore a relatively homogeneous population ($\sigma_u = 10$) and a relatively heterogeneous population ($\sigma_u = 30$). These values were chosen because they are comparable to those used elsewhere in the literature, because they result in moderate-to-low truncation, and because their range illustrates how the accuracy of the methods varies with heterogeneity.

To simulate the response to each metric paired-comparison (PC) question, we calculate the true utility difference between each pair of product profiles by multiplying the design vector by the vector of true partworths: $\vec{x}\vec{u}$. We assume that the respondents' answers to the questions equal the true utility difference plus a zero-mean normal response error with variance σ_{pc}^2. The assumption of normally distributed error is common in the literature and appears to be a reasonable assumption about PC response errors (Wittink and Cattin 1981 report no systematic effects due to the type of error distribution assumed). We select response errors, comparable to those used in the literature. Specifically, to illustrate the range of response errors we use both a low response error ($\sigma_{pc} = 20$) and a high response error ($\sigma_{pc} = 40$).[5] For each comparison, we simulate 500 respondents (in five sets of 100).

[5] Response errors in the literature, often reported as a ratio of error variance to true variance (heterogeneity), vary considerably. In our case, the "respondent's" answer, a, is the difference between the sum of the u_fs. Thus the variance of a is a multiple of σ_u^2. For our situation, the percent errors vary from 8% to 57%.

Question Design Benchmarks

We compare the polyhedral question-design method against three benchmarks: random question design, a fixed design, and the question design used by adaptive conjoint analysis (ACA). For levels within random benchmark, the feature levels are chosen randomly and equally-likely. The fixed design provides another nonadaptive benchmark. For $q > p$, we select the ($q = 16$) design with an algorithm that seeks the highest obtainable D-efficiency (Kuhfield et al. 1994). Efficiency is not defined for $q < p$, thus, for $q = 8$, we follow the procedure established by Lenk et al. (1996) and choose questions randomly from an efficient design for $q = 16$.

We choose ACA question design as our third benchmark because it is the industry and academic standard for within-respondent adaptive question design. For example, Green et al. (1991, p. 215) stated that "in the short span of five years, Sawtooth Software's Adaptive Conjoint Analysis has become one of the industry's most popular software packages for collecting and analyzing conjoint data," and go on to cite a number of academic papers on ACA. Although accuracy claims vary, ACA appears to predict reasonably well in many situations (Johnson 1991, Orme 1999).

The ACA method includes five sections: an unacceptability task (that is often skipped), a ranking of levels within features, a series of self-explicated (SE) questions, the metric paired-comparison (PC) questions, and purchase intentions for calibration concepts. The question design procedure has not changed since it was "originally programmed for the Apple II computer in the late 70s" (Orme and King 2002). It adapts the PC questions based on intermediate estimates (after each question) of the partworths. These intermediate estimates are based on an OLS regression using the SE and PC responses and ensure that the pairs of profiles are nearly equal in estimated utility (utility balance). Additional constraints restrict the overall design to be nearly orthogonal (features and levels are presented independently) and balanced (features and levels appear with near equal frequency).

To avoid handicapping the ACA question design in the initial simulations, we simulate the SE responses without adding error. In particular, Sawtooth Software asks for SE responses using a four-point scale, in which the respondent states the relative importance of improving the product from one feature level to another (e.g., adding automatic film ejection to an instant camera). We set the SE responses equal to the true partworths but discretize the answer to match the ACA scale.

Our code was written using Sawtooth Software's documentation together with e-mail interactions with the company's representatives. We then confirmed the accuracy of the code by asking Sawtooth Software to re-estimate partworths for a small sample of data.

Estimation Benchmark

The two estimation methods are the analytic center (AC) method described earlier and hierarchical Bayes (HB) estimation. Hierarchical Bayes estimation uses data from the population to inform the distribution of partworths across respondents and, in doing so, estimates the posterior mean of respondent-level partworths with an algorithm based on Gibbs sampling and the Metropolis Hastings algorithm (Allenby and Rossi 1999, Arora et al. 1998, Johnson 1999, Lenk et al. 1996, Liechty et al. 2001, Sawtooth Software 2001, Yang et al. 2002). For ACA question design, Sawtooth Software recommends HB as their most accurate estimation method (Sawtooth Software 2002, p. 11). For this initial comparison, for all question-design methods, we use data from the SEs as starting values and we use the SEs to constrain the rank order of the levels for each feature (Sawtooth Software 2001, p. 13).[6]

Criterion

To compare the performance of each benchmark we calculate the mean absolute accuracy of the parameter estimates (true versus estimated values averaged across parameters and respondents). We chose

[6] Another version of Sawtooth Software's HB algorithm also uses the SEs to constraint the relative partworths across features. We test this version in our next set of simulations. This enables us to isolate the impact of the paired-comparison question design algorithm.

to report mean absolute error (MAE) rather than root mean squared error (RMSE) because the former is less sensitive to outliers and is more robust over a variety of induced error distributions (Hoaglin et al. 1983, Tukey 1960). However, as a practical matter, the qualitative implications of our simulations are the same for both error measures. Indeed, except for a scale change, the results are almost identical for both MAE and RMSE. This is not surprising; for normal distributions the two measures differ only by a factor of $(2/\pi)^{1/2}$.

The results are based on the average of five simulations, each with 100 respondents. To reduce unnecessary variance among question-design methods, we first draw the partworths and then use the same partworths to evaluate each question-design method. The use of multiple draws makes the results less sensitive to spurious effects from a single draw.

4. Results of the Initial Monte Carlo Experiments

We begin with the results obtained from using eight ($q = 8$) paired comparison questions. This is the type of domain for which polyhedral question design and analytic center estimation were developed (more parameters to estimate than there are questions). Moreover, within this domain there are generally a range of partworths that are feasible, so the polyhedron is not empty. In our simulations the polyhedron contains feasible answers for an average of 7.97 questions when response errors are low. When response errors are high, this average drops to 6.64.

Table 1 reports the MAE in the estimated partworths for a complete crossing of question-design methods, estimation methods, response error, and heterogeneity. The best results (lowest error) in each column are indicated by **bold text**. In Table 2 we reorganize the data to indicate the directional impact of either heterogeneity or response errors on the performance of the question-design and estimation methods. In particular, we average the performance of each question-design method across estimation methods (and vice versa). To indicate the directional effect of heterogeneity we average across response errors (and vice versa).

Question-Design Methods

The findings indicate that when there are only a small number of PC questions, the polyhedral question-design method performs well compared to the other three benchmarks. This conclusion holds across the different levels of response error and heterogeneity. The improvement over the random question-design method is reassuring, but perhaps not surprising. The improvement over the fixed method is also not surprising when there are a small number of questions, because it is not possible to achieve the balance and orthogonality goals that the fixed method seeks.

Table 1 Comparison of Question-Design and Estimation Methods for $q = 8$, Mean Absolute Errors

		Homogeneous Population		Heterogeneous Population	
Question Design	Estimation	Low Response Error	High Response Error	Low Response Error	High Response Error
Random	AC	16.5	24.1	15.9	21.7
	HB	8.1	10.2	19.8	22.2
Efficient fixed	AC	13.7	22.9	14.3	21.0
	HB	**7.8***	10.3	20.4	22.5
ACA	AC	14.9	24.2	16.1	22.1
	HB	8.3	**9.8***	23.9	22.9
Polyhedral	AC	10.7	20.9	**12.5***	**19.7***
	HB	**7.8***	**9.9***	20.6	22.2

Notes. Smaller numbers indicate better performance.

*For each column, lowest error or not significantly different from lowest ($p < 0.05$). All others are significantly different from lowest.

Table 2 Directional Implications of Response Errors and Heterogeneity for $q = 8$, Mean Absolute Errors

	Homogeneous Population	Heterogeneous Population	Low Response Error	Low Response Error
Question design				
Random	14.7	19.9	15.1	19.5
Efficient fixed	13.6	19.5	14.0	19.1
ACA	14.3	21.2	15.8	19.8
Polyhedral	**12.3***	**18.4***	**12.9***	**18.2***
Estimation				
AC	18.5	**17.9***	**14.3***	22.1
HB	**9.0***	21.8	**14.6***	**16.2***

Notes. Smaller numbers indicate better performance.

*For each column within question design or estimation, lowest error or not significantly different from lowest ($p < 0.05$). All others are significantly different from lowest.

The comparison with ACA question design is more interesting. Further investigation reveals that the relatively poor performance of the ACA method can be attributed, in part, to endogeneity bias, resulting from utility balance—the method that ACA uses to adapt questions. To understand this result we first recognize that any adaptive question-design method is potentially subject to endogeneity bias. Specifically, the qth question depends upon the answers to the first $q-1$ questions. This means that the qth question depends, in part, on any response errors in the first $q-1$ questions. This is a classical problem, which often leads to bias (see, for example, Judge et al. 1985, p. 571). Thus, adaptivity represents a trade-off: We get better estimates more quickly, but with the risk of endogeneity bias. In our simulations, the absolute bias with ACA questions and AC estimation is approximately 6.6% of the mean when averaged across domains. This is statistically significant, in part because of the large sample size in the simulations. Polyhedral question design is also adaptive and it, too, could lead to biases. However, in all four domains, the bias for ACA questions is significantly larger than the bias for polyhedral questions (1.0% on average for AC). The endogeneity bias in ACA questions appears to be from utility-balanced question design; it is not removed with HB estimation. Detailed results are available from the authors.[7] While further analyses of endogeneity bias are beyond the scope of this paper, they represent an interesting topic for future research. In particular, it might be possible to derive estimation methods that correct for these endogeneity biases.

Estimation Methods

For homogeneous populations, hierarchical Bayes consistently performed better than analytic center estimation, irrespective of the question-design method. The performance differences were generally large. Hierarchical Bayes estimation uses population-level data to moderate individual estimates. If the population is homogenous, then at the individual level, the ratio of noise to true variation is higher, so moderating this variance through population-level data improves accuracy. However, if the population is heterogeneous, then reliance on population data makes it more difficult to identify the true individual-level variation, and analytic center estimation does better. For a heterogeneous population, the combination of polyhedral question design and analytic center estimation was significantly more accurate than any other combination of question-design or estimation method (Table 1).

The findings also suggest that hierarchical Bayes is relatively more accurate when response errors are high, while analytic center estimation is more likely to be favored when response errors are low. The reliance of hierarchical Bayes on population-level data may also explain the role of response errors. If response errors are large, much of the individual-level variance is due to noise. Population-level data are less sensitive to response errors (due to aggregation), so reliance on this data helps to improve accuracy. On the other hand, when response errors are low, the polyhedron stays feasible longer and the analytic center method appears to do a better job of identifying individual-level variation.

[7] The endogeneity biases persist in most domains for $q = 16$. In most cases the endogeneity biases for ACA questions are larger than those for polyhedral questions. We have been able to show formally that utility balance leads to bias for OLS, but we have not yet been able to construct proofs for AC or HB. Proofs available from the authors.

Additional Paired-Comparison Questions

Although polyhedral methods were developed primarily for situations with only a relatively small number of questions, there remain important applications in which a larger number of questions can be asked of each respondent. To examine whether the potential accuracy advantages of polyhedral methods for low q leads to a loss of accuracy at high q, we re-examined the performance of each method after 16 paired-comparison questions ($q = 16$).

Recall that the polyhedral method is used to design questions only when the polyhedron contains feasible responses. For low response errors the polyhedron is typically empty after eight questions, while for high response errors this generally occurs at around six or seven questions. Once the polyhedron is empty, we choose questions randomly. Because the polyhedral question-design method is only responsible for around half of the questions we use the label "poly/random."

The findings are reported in Table 3, which is analogous to Table 2. They reveal the emergence of fixed question-design methods in some domains. Asking a larger number of questions results in more complete coverage of the parameter space. This increases the importance of orthogonality and balance—the criteria used in efficient fixed question design. With more complete coverage, the ability to customize questions to focus on specific regions of the question space becomes less important, mitigating the advantage offered by adaptive techniques.

Table 3 **Directional Implications of Response Errors and Heterogeneity for $q = 16$, Mean Absolute Errors**

	Homogeneous Population	Heterogeneous Population	Low Response Error	High Response Error
Question design				
Random	12.1	14.3	10.4	15.9
Efficient fixed	10.3	**13.1***	**8.8***	14.6
ACA	12.5	18.1	13.5	17.2
Poly/random	**9.2***	15.2	10.4	**14.0***
Estimation method				
AC	13.9	**12.5***	**9.9***	16.4
HB	**8.2***	17.9	11.6	**14.4***

Notes. Smaller numbers indicate better performance.

*For each column within question design or estimation, lowest error or not significantly different from lowest ($p < 0.05$). All others are significantly different from the lowest.

However, even after 16 questions there remain domains in which polyhedral question design can improve performance. The poly/random method appears to be at least as accurate as the fixed design when the population is homogenous and/or response errors are high. Its advantage for low q does not seem to be particularly harmful for high q, especially for high response errors. Table 3 also suggests that analytic center estimation remains a useful estimation procedure when populations are heterogeneous and/or response errors are low. We note that this result is consistent with Andrews et al. (2002, p. 87), who conclude "individual-level models overfit the data." They test OLS rather than the analytic center method and do not test adaptive methods.

In summary, our Monte Carlo experiments suggest that there are domains in which polyhedral question design and/or analytic center estimation improve the accuracy of conjoint analysis, but there are also domains better served by extant methods. Specifically,

- Polyhedral question design shows promise for low number of questions, such as the fuzzy front-end of product development and/or Web-based interviewing.
- For larger numbers of questions, efficient fixed designs appear to be best, but poly/random question design does well, especially when response errors are high and populations are homogenous.
- Analytic center estimation shows promise for heterogeneous populations and/or low response errors where the advantage of an individual-respondent focus is strongest.
- Hierarchical Bayes estimation is preferred when populations are more homogeneous and response errors are large.

5. The Role of Self-Explicated Questions

Hybrid conjoint models refer to methods that combine both compositional methods, such as self-explicated (SE) questions, and decompositional methods, such as metric paired-comparison (PC) questions, to produce new estimates. Although there are instances in

which methods that use just one of these data sources outperform or provide equivalent accuracy to hybrid methods, there are many situations and product categories in which hybrid methods improve accuracy (e.g., Green 1984).

An important hybrid from the perspective of evaluating polyhedral methods is ACA—the most widely used method for adaptive metric paired-comparison questions. While ACA's question design algorithm has remained constant since the late 1970s, its estimation procedures have evolved to address the incommensurability of the SE and PC scales. Its default estimation procedure relies on an ordinary least-squares (OLS) regression that weighs the SE and the PC data in proportion to the number of questions asked (Sawtooth Software 2002, Version 5).[8] We label the current version "weighted hybrid" estimation and denote it by the acronym WHSE (the SE suffix indicates reliance on SE data). Sawtooth Software also incorporates the SE responses in their hierarchical Bayes estimation procedure by using the SEs to constrain the estimates of partworths both within a feature and between features to satisfy the ordinal conditions imposed by the SE data. For example, if a respondent's responses to the SE questions indicate that picture quality is more important than battery life, then the hierarchical Bayes parameters are restricted to satisfying this condition. We denote this algorithm with the acronym HBSE to indicate that the SE responses play a larger role in the estimation.

We also create a polyhedral hybrid by extending AC estimation to incorporate SE responses. To do so, we introduce constraints on the feasible polyhedron similar to those used by HBSE. For example, we impose a condition that picture quality is more important than battery life by using an inequality constraint on the polyhedron to exclude points in the partworth space that breach this condition. When the polyhedron becomes empty, we extend OPT4 to incorporate both the PC and SE constraints. We distinguish this method from the analytic center method by adding a suffix to the acronym: ACSE.

[8] Earlier versions of ACA either weighed the scales equally (Version 3) or selected weights to fit purchase-intention questions (Version 4).

To compare WHSE, HBSE, and ACSE to their pure-bred progenitors, we must consider the accuracy of the SE data. If the SE data are perfectly accurate, then a model based on SEs alone will predict perfectly, and the hybrids would be almost as accurate. On the other hand, if the SEs are extremely noisy, then the hybrids may actually predict worse than methods that do not use SE data. To examine these questions, we undertook a second set of simulation experiments.

To simulate SE responses we assume that respondents' answers to SE questions are unbiased but imprecise. In particular, we simulate response error in the SE answers by adding to the vector of true partworths, $\vec{u}$, a vector of independent identically distributed normal error terms with variance σ_{se}^2. We simulate two levels of SE response error—low error relative to PC responses ($\sigma_{se} = 10$) and high error relative to PC responses ($\sigma_{se} = 70$), truncating the SEs to a 0-to-100 scale.[9] We expect that these benchmarks should bound empirical situations. Recall that in the first set of simulations we assumed no SE errors ($\sigma_{se} = 0$), but discretized the scale. For consistency, we also use a discrete scale when there are nonzero SE errors. Based on these SE errors, we redo the simulations for each level of PC response errors and heterogeneity.

We summarize the results with Table 4 for lower numbers of questions ($q = 8$), where we report the most accurate question-design/estimation methods for each level of response error and heterogeneity. Detailed results are available from the authors. For ease of comparison, the earlier results (from Table 1) are summarized in the column labeled "Initial Simulations."

There are three results of interest. First, when the SEs are more accurate than the PCs, then the hybrids do well. In this situation, the PC question-design method matters less: polyhedral, fixed, and random hybrids are not significantly different in accuracy. Second, the insights obtained from Tables 1 through 3

[9] For high SE errors truncation is approximately 50%; for low SE errors truncation is approximately 0% and 11%, respectively, for low and high heterogeneity. This is consistent with our manipulation of high versus low information content of the SEs.

Table 4 The Impact of Self-Explicated (SE) Questions ($q = 8$)

Heterogeneity	Response Errors	Initial Simulations (No SEs)	Relatively Accurate SEs	Relatively Noisy SEs
Homogeneous	Low error	Polyhedral HB Fixed HB	Polyhedral HBSE Fixed HBSE Random HBSE	Polyhedral HB Fixed HB
	High error	Polyhedral HB ACA HB	Polyhedral HBSE Fixed HBSE Random HBSE	Polyhedral HBSE
Heterogeneous	Low error	Polyhedral AC	Polyhedral WHSE Fixed WHSE Random WHSE	Polyhedral AC
	High error	Polyhedral AC	Polyhedral WHSE Fixed WHSE Random WHSE	Polyhedral AC

for population-level versus individual-level estimation continue to hold: HB or HBSE do well in homogeneous domains while AC or WHSE do well in heterogeneous domains. Third, when the SEs are noisy relative to the PCs, then the hybrid methods do not do as well as the purebred methods. Indeed, we expect a crossover point at some intermediate level of relative accuracy.

Table 4 also highlights the emergence of WHSE in some domains.[10] Of the hybrids tested, WHSE is the only method that makes use of the interval-scale properties of the SEs. These metric properties appear to help when the "signal-to-noise ratio" is high (more variation in true partworths, less error in the SEs). This result suggests that other methods which use interval-scaled properties of the SEs should do well in these domains—a topic for further hybrid development (e.g., Ter Hofstede et al. 2002).

In summary, as in biology, where genetically diverse offspring often have traits superior to their purebred parents, heterosis in conjoint analysis improves predictive accuracy in some domains. Furthermore, polyhedral question design remains promising in these domains, and many of the insights from our earlier simulations still hold for hybrid methods. Finally, we expect and obtain analogous results for larger numbers of questions ($q = 16$). We could identify no additional insight beyond Tables 1 through 4.

[10] If WHSE were not available, polyhedral ACSE is best for low response errors and polyhedral HBSE is best for high response errors. These results suggest that there is room for future research on how best to incorporate SE data with AC estimation.

6. Empirical Application and Test of Polyhedral Methods

While tests of internal validity are common in the conjoint-analysis literature, tests of external validity at the individual level are rare.[11] 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 student subjects, respectively, completed a conjoint questionnaire prior to accepting job offers. The methods were compared on their ability to predict actual job choices. First-preference predictions ranged from 64% to 76% versus random-choice percentages of 26 to 36%. 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 the applications when families were prompted to

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

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 et al. (1984) allocated 122 undergraduate business majors randomly to 12 different conjoint tasks designed to measure partworths for 5 features. Respondents indicated their preferences for 10 calculators offered in a 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 this section, we test polyhedral methods with an empirical test involving an innovative laptop-computer carrying bag. Our test differs from the natural experiment studies because it is based on a controlled experiment in which we chose pareto sets of product features. At the time of our study, the product was not yet on the market, so respondents had no prior experience with it. 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 feature. Price is restricted to two levels ($70 and $100)—the extreme prices for the bags in both the internal and external validity tests. We estimated the partworths associated with prices between $70 and $100 by linearly interpolating. A more detailed description of the product features can be found on the Web sites cited earlier in the paper (and in the Acknowledgments).

Because ACA is the dominant industry method for adaptive question design, we chose a product category where we expected ACA to perform well—a category where separable product features would lead to moderately accurate SE responses. We anticipate that SE responses are more accurate in categories where customers make purchasing decisions about features separately 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. If polyhedral question design and/or estimation does well in this category, then, based on Table 4, we expect it to do well in categories where SE responses are less accurate.

Research Design

Subjects were randomly assigned to one of the three conjoint question-design methods: polyhedral (two cells), fixed, or ACA. We omitted random question design because the fixed question-design method dominates random design in Tables 2 and 3. After completing the respective conjoint tasks, all the respondents were presented with the same validation exercises. The internal validation exercise involved four holdout metric paired-comparison (PC) questions, which occurred immediately after the 16 PC questions designed by the respective conjoint methods. The external validation exercise was the selection of a laptop-computer bag from a choice set of five bags. This exercise occurred in the same session as the conjoint tasks and holdout questions, but was separated from these activities by a filler task designed to cleanse memory (see Table 5).

Conjoint Tasks

Recall that ACA requires five sets of questions. Pretests confirmed that all of the features were acceptable to the target market, allowing us to skip the unacceptability task. This left four remaining tasks: ranking of levels within features, self-explicated (SE) questions, metric paired-comparison (PC) questions, and purchase intention (PI) questions. ACA

Table 5 Detailed Research Design

Row	Polyhedral 1	Fixed	Polyhedral 2	ACA
1			Self-explicated	Self-explicated
2	Polyhedral paired comparison	Fixed paired comparison	Polyhedral paired comparison	ACA paired comparison
3	Internal validity task	Internal validity task	Internal validity task	Internal validity task
4			Purchase intentions	Purchase intentions
5	Filler task	Filler task	Filler task	Filler task
6	External validity task	External validity task	External validity task	External validity task

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 the ACA condition.

The fixed and polyhedral question design techniques do not require SE or PI questions. Because asking the SE questions first could create a question-order effect, we asked only PC questions (not the SE or PI questions) prior to the validation task in the fixed condition. To investigate the question-order effect we included two polyhedral data collection procedures: one that matched the fixed design (polyhedral 1) and one that matched ACA (polyhedral 2). In polyhedral 1, only PC questions preceded the validation task; while in polyhedral 2, all the questions preceded the validation task. This enables us to (a) explore whether the SE questions affect the responses to the PC questions and (b) evaluate the hybrid estimation methods that combine data from PC and SE questions.[12]

The complete research design, including the question order, is summarized in Table 5. Questions associated with the conjoint tasks are highlighted in green (rows 1, 2, and 4), while the validation tasks are highlighted in yellow (rows 3 and 6). The filler task is highlighted in blue (row 5). In this design, polyhedral 1 can be matched with fixed; polyhedral 2 can be matched with ACA.

Internal Validity Task: Holdout PC Questions

Each of the question-design methods designed 16 metric paired-comparison (PC) questions. Following these questions, respondents answered four holdout PC questions—a test used extensively in the literature. The holdout profiles were randomly selected from an independent efficient design of 16 profiles and did not depend on prior answers by that respondent. There was no separation between the 16 initial questions and the four holdout questions, so that respondents were not aware that the questions were serving a different role.

[12] Although the SE responses are collected in the polyhedral 2 condition, they are not used in analytic center estimation or polyhedral question design. However, they do provide the opportunity to test hybrid estimation methods.

Filler Task

The filler task was designed to separate the conjoint tasks and the external validity task. It was hoped that this separation would mitigate any memory effects that might influence how accurately the information from the conjoint tasks predicted which bags respondents chose in the external validity tasks. 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 the responses to the filler task across the four conditions.

External Validity Task: Final Bag Selection

Respondents were told that they had \$100 to spend and were asked to choose among five bags. The five bags shown to each respondent were drawn randomly from an orthogonal fractional factorial design of 16 bags. This design was the same across all four experimental conditions, so that there was no difference, on average, in the bags shown to respondents in each condition. 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 anticipated market price of the features included with each bag. By pricing the bags in this manner we ensured that the choice set represented a Pareto frontier, as recommended by Elrod et al. (1992), Green et al. (1988), and Johnson et al. (1989).

Respondents were instructed that they would receive the bag they chose. If the bag was priced at less than \$100, they were promised cash for the difference. 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 6.[13] At the end of the study the chosen bags were distributed to respondents together with

[13] We acknowledge two trade-offs in this design. The 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, a priori neither should favor one of the three methods relative to the other; we expect no interaction between the endowment/forced-

Figure 6 Respondents Choose and Keep a Laptop Computer Bag

the cash difference (if any) between the price of the selected bag and $100.

Self-Explicated and Purchase Intention Questions

The self-explicated questions asked respondents to rate the importance of each of the 10 product features. For a fair comparison to ACA, we used the wording for the questions, the (four-point) response scale, and the algorithm for profile selection proposed by Sawtooth Software (1996). For the purchase intention questions, respondents were shown six bags, and we asked how likely they were to purchase each bag. We adopted the wording, response scale, and algorithms for profile selection suggested by Sawtooth Software.

Subjects

The subjects (respondents) were first-year MBA students. They were not informed about the objectives of the study, nor had they taken a course in which conjoint analysis was taught in detail. We received 330 complete responses (there was one incomplete response) from an e-mail invitation to 360 students—a response rate of over 91%. Pure random assignment (without quotas) yielded 80 subjects for the ACA condition, 88 for the fixed condition, and 162 for the polyhedral conditions, broken out as 88 for the standard question order (polyhedral 1) and 74 for the alternative question order (polyhedral 2).

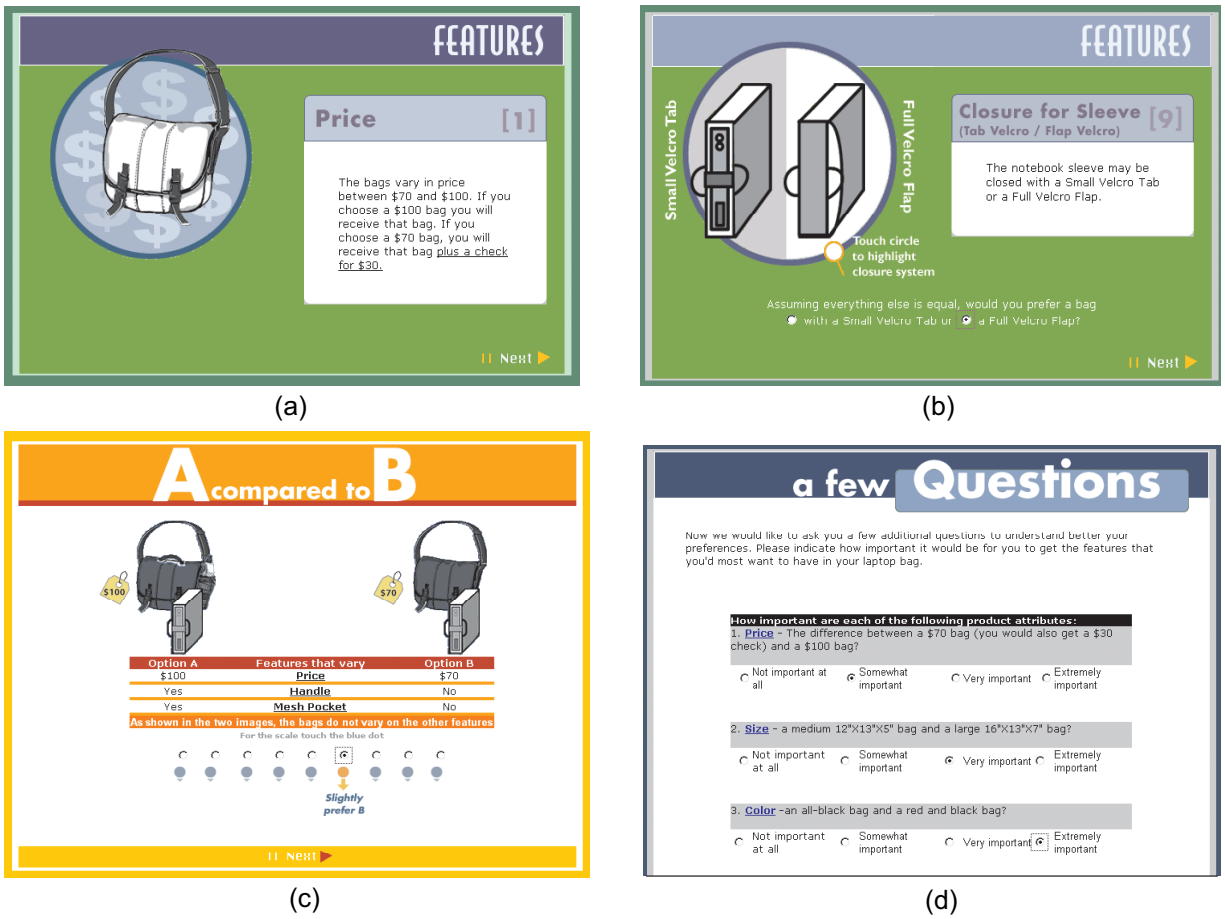
The questionnaires were pretested on a total of 69 subjects drawn from professional market research and consulting firms, former students, graduate students in operations research, and second-year students in an advanced marketing course that studied conjoint analysis. The pretests were valuable for 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.

Additional Details

Figure 7 illustrates some of the key screens in the conjoint analysis questionnaires. In Figure 7a respondents are introduced to the price feature. Figure 7b illustrates one of the dichotomous features—the closure on the sleeve. This is an animated screen that provides more detail as respondents move their pointing devices past the picture. Figure 7c illustrates one of the PC tasks. Respondents were asked to rate their relative preference for two profiles that varied on three features. Both text and pictures were used to describe the profiles. In the pictures, features that did not vary between the products were chosen to coincide with the respondent's preferences for feature levels obtained in the tasks such as Figure 7b. The format was identical for all four experimental treatments. Finally, Figure 7d illustrates the first three self-explicated questions. The full questionnaires for each treatment are available on a Web site cited earlier in this paper. We note that some of these Web site improvements (e.g., dynamically changing

choice design and PC question design and leave such investigations to future research. However, the forced choice design might add noise to the most accurate methods relative to less accurate methods. This would make it more difficult to achieve significant differences and is thus conservative. Pragmatically, we designed the task to maximize the power of the statistical comparisons of the four treatments. The forced choice also helped to reduce the (substantial) cost of this research.

Figure 7 Example Screens from Questionnaires



Notes. (a) Price as change from $100. (b) Introduction of "Sleeve" feature. (c) Metric paired-comparison (PC) question. (d) Self-explicated (SE) questions.

pictures) are not standard in Sawtooth Software's implementation; thus, our tests should be considered a test of ACA question design (and estimation) rather than a test of Sawtooth Software's commercial implementation.

7. Results of the Field Test

To evaluate the conjoint methods we calculated the Spearman rank-order correlation between the actual and observed rankings for the five bags shown to each respondent.[14] We report the results in Table 6 using the same benchmark methods that we used for the Monte Carlo simulations.

In the simulation analysis we had the luxury of large sample sizes (500 respondents) and were able to completely control for respondent heterogeneity. Although the sample sizes in Table 6 are large compared to previous tests of this type, they are small compared to the simulation analysis. As a result, none of the differences across methods are significant at the 0.05 level in independent-sample *t*-tests. However,

[14] As an alternative metric, we compared how well the methods predicted which product the respondents favored. The two metrics provide a similar pattern of results so, for ease of exposition, we focus on the correlation measure. There are additional reasons to focus on correlations. First-choice prediction is a dichotomous variable highly dependent upon the number of items in the choice set. In addition, it provides less power because it has higher variance than the Spearman correlation, which is based on a rank order of five items.

Table 6 External Validity Tests: Correlation with Actual Choice

	After 8 Questions		After 16 Questions	
Methods Without SE Data	Fixed Questions	Polyhedral 1 Questions	Fixed Questions	Polyhedral 1 Questions
Analytic center (AC)	0.51	0.59	0.62	0.68
Hierarchical Bayes (HB)	0.53	0.57	0.61	0.64
Sample size	88	88	88	88
Methods That Use SE Data	ACA Questions	Polyhedral 2 Questions	ACA Questions	Polyhedral 2 Questions
WHSE estimation	0.66	0.68	0.68	0.72
ACSE estimation	0.63	0.70	0.65	0.71
HBSE estimation	0.64	0.73	0.65	0.74
Sample size	80	74	80	74

Note. Larger numbers indicate better performance.

these independent-sample t-tests do not use all the information available in the data. We also evaluate significance by an alternative method that pools the correlation measures calculated after each additional PC question. This results in a total of sixteen observations for each respondent.

To control for heteroscedasticity we estimate a separate intercept for each question number. We also controlled for respondent heterogeneity in the randomly assigned conditions with a null model that assumes that the 10 laptop bag features are equally important. If, despite the random assignment of respondents to conditions, the responses in one condition are more consistent with the null model, then the comparisons would be biased in favor of this condition. We control for such potential heterogeneity by including a measure describing how accurately the equal-weights (null) model performs on the predictive correlations. The complete specification for this model is described in Equation (1), where r indexes the respondents and q indexes the number of PC questions used in the partworth estimates. The αs and βs are coefficients in the regression and ε_{rq} is an error term.

$$Correlation_{rq} = \sum_{q=1}^{16} \alpha_q \, Question_q + \sum_{m=1}^{M-1} \beta_m \, Method_m + \gamma \, EqualWeights_r + \varepsilon_{rq}. \quad (1)$$

The *Question* and *Method* terms refer to dummy variables identifying the question and method effects. The *EqualWeight* variable measures the correlation obtained for respondent r between the actual rankings and the rankings obtained from an equal weights model. Under this specification, the β coefficients represent the expected increase or decrease in this correlation across questions due to method m relative to an arbitrarily chosen base method. Positive (negative) values for the β coefficients indicate that the correlations between the actual and predicted rankings are higher (lower) for method m than the base method.

We further control for potential heteroscedasticity introduced by the panel nature of the data by calculating robust standard errors (White 1980). We also estimated a random effects model, but there were almost no differences in the coefficients of interest. Moreover, the Hausman specification test favored the fixed-effects specification. The findings are summarized in Table 7.

Comparison of Estimation Methods

We compare the accuracy of the different estimation methods by comparing the findings in Table 6 within a column (for a specific set of questions) and looking to Table 7 for significance tests. This comparison holds the question design constant and varies the estimation method. For those experimental cells that were designed to obtain estimates without the SE questions, hierarchical Bayes and analytic center estimation offer similar predictive accuracy for fixed

Table 7 External Validity Tests: Conclusions from the Multivariate Analysis

	Without SE Questions	With SE Questions
Comparison of Estimation Methods		
Fixed questions	HB > AC	
Polyhedral 1 questions	AC ≫ HB	
ACA questions		WHSE > HBSE > ACSE
Polyhedral 2 questions		HBSE ⋙ WHSE > ACSE
Comparison of Question-Design Methods		
AC estimation	Polyhedral 1 ⋙ fixed	
HB estimation	Polyhedral 1 ⋙ fixed	
WHSE estimation		Polyhedral 2 > ACA
ACSE estimation		Polyhedral 2 ≫ ACA
HBSE estimation		Polyhedral 2 ⋙ ACA

Notes. Method m > Method n: Method *m* is more accurate than method *n* but the difference is not significant.

Method m ≫ Method n: Method *m* is significantly more accurate than method *n* ($p < 0.05$).

Method m ⋙ Method n: Method *m* is significantly more accurate than method *n* ($p < 0.01$).

questions, but analytic center estimation performs better for polyhedral questions.

If SE responses are available, the preferred estimation method appears to depend upon both the question-design method and the number of PC responses used in the estimation. For polyhedral questions HBSE performs extremely well for low numbers of PC questions, perhaps due to its use of population level data. However, increasing the number of PC responses yields less improvement in the accuracy of HBSE relative to WHSE. After 16 questions all three estimation methods converge to comparable accuracy levels, suggesting that there are sufficient data at the individual level to provide estimates that need not depend on population distributions. When using PC questions designed by ACA, WHSE out-performs HBSE, albeit not significantly so.

Comparison of Question-Design Methods

The findings in Table 6 also facilitate comparison of the question-design methods. Comparing across columns (within rows) in Table 6 holds the estimation method constant and varies the question design. The findings favor the two conditions in which the polyhedral question design was used. When the SE measures were not collected, the polyhedral question design yielded significantly ($p < 0.01$) more accurate predictions than the fixed design. This holds true irrespective of the estimation method.

When SE responses were collected, the polyhedral question design was more accurate than ACA across every estimation method, although the difference was not significant for WHSE. Detailed investigation reveals that for every estimation method we tested, the estimates derived using the polyhedral questions outperform the corresponding estimates derived using ACA questions after each and every question number.

The Incremental Predictive Value of the SE Questions and of PC Questions

In this category, Table 6 suggests that hybrid methods that use both SE and PC questions consistently outperform methods that rely on PC questions alone. Thus, in this category, the SE questions provide incremental predictive ability. We caution that the product category was chosen at least in part because the SE responses were expected to be accurate. The simulations suggest that this improvement in accuracy may not be true in all domains.

We also evaluate whether the PC responses contributed incremental accuracy. Predictions that use the SE responses alone (without the PC responses) yield an average correlation with actual choice of 0.64. This is lower than the performance of the best methods that use both SE and PC responses and comparable at $q = 16$ to those methods that do not use SE responses (see Table 6). We conclude (in this category) that the

16 PC questions provide roughly the same amount of information as the 10 SE questions and that, for methods that use both, the PC data add incremental predictive ability. This conclusion is consistent with previous evidence in the literature (Green et al. 1981, Huber et al. 1993, Johnson 1999, Leigh et al. 1984).

The Internal Validity Task

We repeated the analysis of question design and estimation methods using the correlation measures from the internal validity (holdout questions) task. Details are in Appendix 2. The results for internal validity are similar to the results for external validity. However, there are two differences worth noting. First, while HBSE predicted better than WHSE for polyhedral question design in the choice task, there was no significant difference in the holdout task. Second, while polyhedral question design was significantly better than fixed design for the choice task, there were no significant differences for the holdout task.

Question Order Effects: Polyhedral 1 Versus Polyhedral 2

Polyhedral 1 and polyhedral 2 varied in question order; the SE questions preceded the PC questions in polyhedral 2 but not in polyhedral 1. Otherwise, both methods used the same question design algorithm—an algorithm that does not use SE data. Nonetheless, question order might influence the accuracy of the PC responses. If the SE questions "wear out" or tire respondents, causing them to pay less attention to the PC questions, we might expect that inclusion of the SE questions will degrade the accuracy of the PC responses. Alternatively, the SE questions may improve the accuracy of the PC questions by acting as a training or "warm-up" task that helps respondents clarify their values, increasing the accuracy of the PC questions (Green et al. 1991, Huber et al. 1993, Johnson 1991).

By comparing the two experimental cells we investigate whether the prior SE questions affected the accuracy of the respondents' PC responses. The predictive accuracy of the two conditions are not statistically different ($t = -0.05$ for AC estimation, the preferred estimation method from Tables 1 and 7). This suggests that by the sixteenth question any wear out or warm-up/learning had disappeared. However, there might still be an effect for the early questions. When we estimate performance of AC estimation using a version of Equation (1), the effect is not significant for external validity task ($t = 0.68$) but is significant for the internal validity task ($t = 2.60$). In summary, the evidence is mixed. There is no evidence that the SE questions improve or degrade the accuracy of the PC questions for the choice task, but they might improve accuracy for the holdout task. Further testing is warranted. For example, if the first cut of the polyhedron is critical, then it is important that the first PC question be answered as accurately as feasible. The use of SE questions might sensitize the respondent and enhance the accuracy of the first PC question. Alternatively, researchers might investigate other warm-up questions, such as those in which the respondent configures an ideal product (Dahan and Hauser 2002).

Summary of the Field Test

In the field test, polyhedral question-design appears to be the most accurate of the tested question-design methods. When SE data are available, the most accurate estimation methods were the HBSE and WHSE hybrids. If SE data were unavailable, the most accurate estimation method was AC for polyhedral questions and HB for fixed questions.

To compare the field test to the simulations, we must identify the relevant domain. Fortunately, estimates of heterogeneity and PC response errors are a by-product of the hierarchical Bayes estimation, and we can use HB to estimate SE errors. These estimates suggest high levels of heterogeneity ($\sigma_u^2 \approx 29$) and PC response errors ($\sigma_{pc}^2 \approx 43$) but moderately low SE response errors ($\sigma_{pc}^2 \approx 18$). When SEs are unavailable, the simulations predict that in this domain: (1) Polyhedral question design should be better than fixed for both estimation methods, (2) AC should be much better than HB for polyhedral questions, (3) AC should remain better than HB for fixed questions, but the difference is not as large. The significant findings in Table 7 are consistent with (1) and (2). Contrary to (3), HB is better for fixed questions, but not significantly so.

For accurate SEs, the simulations predict that in this domain: (4) Polyhedral questions will remain strong for hybrid estimation methods, but the differences among question-design methods will be less for hybrid methods than for purebred methods; (5) hybrid estimation methods will outperform the purebred methods; and (6) WHSE will outperform ACSE (in the detailed simulation data for this domain HBSE also outperforms ACSE). Predictions (4), (5), and (6) hold true in the field test.

It is always difficult to compare field data to simulations because, despite experimental controls, there may be unobserved phenomena in the field test that are not captured in the simulations. However, the two types of data are remarkably consistent, albeit not perfectly so.

8. Product Launch

Subsequent to our research, Timbuk2 launched the laptop bags with features similar to those tested including multiple sizes, custom colors, logo options, accessory holders (PDA and cellular phone), mesh pockets, and laptop sleeves. Timbuk2 considers the product a success—It is selling well and is profitable. We now compare the laboratory experiment and the national launch. However, we do so with caution because the goal of the field test was to compare methods rather than to forecast the national launch. By design we used a student sample rather than a national sample, offered only two color combinations, and did not offer the large size bag. Furthermore, one tested feature, the "boot," was not included in the national launch because production cost (and feasibility) exceeded the price that could be justified. One feature, a bicycle strap, was added based on managerial judgment.

There were five comparable features that appeared in both the field test and the national launch. With the above caveats in mind, the correlation of the predicted feature shares from the conjoint analyses with those observed in the marketplace was 0.9, which was significant. (By feature share we mean percent of customers who chose each of the five features.) Predictions with various null models were not significant. Unfortunately, these data do not provide sufficient power to compare the relative accuracies of the methods nor report correlations to more than one significant figure.

9. Conclusions and Future Research

Recent developments in math programming provide new methods for designing metric paired-comparison questions and estimating partworths. The question-design method uses a multidimensional polyhedron to characterize feasible parameters and focus questions to reduce the size of the polyhedron as fast as possible. The estimation method uses the analytic center to approximate the center of the polyhedron. This centrality estimate summarizes what is known about the feasible set of parameters. Our goals in this paper were to (1) propose practical algorithms using polyhedral methods, (2) demonstrate their feasibility, (3) test their potential in a variety of domains, (4) compare their theoretical accuracy relative to existing methods, and (5) compare their predictive accuracy relative to existing methods in a realistic empirical situation. The field test was designed to match one of the theoretical domains that was favorable to existing methods. The overall conclusion is that polyhedral methods are worth further development, experimentation, and study. Detailed findings are summarized in Table 8.

We are encouraged by the performance of polyhedral methods in the Monte Carlo and external-validity experiments. We feel that with further research new algorithms based on polyhedral methods have the potential to become easier to use, more accurate, and applicable in a broader set of domains. We close by highlighting some of the opportunities.

Theoretical Improvements

Error Theory. Our proposed heuristic (OPT4) provides a practical means by which to use the analytic center to obtain partworth estimates from noisy data. Although the maximum error, $\vec{\delta}$, is likely to increase with the number of questions, the mean error, $\vec{\delta}/q$, is likely to decrease. Furthermore, the analytic center estimates appear to be unbiased, except when there

Table 8 Detailed Summary of Findings

Feasibility
- The polyhedral question-design method can design questions in real time for a realistic number of parameters.
- The analytic center estimation heuristic yields real time partworth estimates that provide reasonable accuracy with little or no bias.

Monte Carlo Experiments
- Polyhedral question design shows the most promise for lower numbers of questions where it does well in all tested heterogeneity and response-error domains.
- Fixed question design remains best for larger numbers of questions, but poly/random does well for homogeneous populations and high response errors.
- AC estimation shows promise for heterogeneous populations and low response errors; HB performs well when the population is homogeneous and response errors are high.
- When SE data are available and relatively noisy:
  - Polyhedral question design continues to perform well.
  - Purebred estimation methods may be preferred.
- When SE data are available and relatively accurate:
  - Question design is relatively less important.
  - For homogeneous populations, HBSE is the best estimation method.
  - For heterogeneous populations, WHSE is the best estimation method.

External-Validity Experiment
- The field test domain had high heterogeneity and PC response errors, and moderately low SE response errors. The findings are consistent with the Monte Carlo results in this domain.
- Polyhedral question design shows promise irrespective of the availability of SE data. This is true for all tested estimation methods.
- When SE data are unavailable, analytic center estimation is a viable alternative to hierarchical Bayes estimation, especially when paired with polyhedral question design.
- When SE data are available, existing hybrid estimation methods (HBSE and WHSE) appear to outperform the ACSE hybrid.

Product Launch
- The laptop bags were launched to the market, but we must be cautious when evaluating predictive accuracy because there were many differences between the empirical experiment and the market launch. In addition, there were insufficient data for relative comparisons.
- With these caveats, the conjoint analyses correlate well with the marketplace outcome.

is endogeneity in question design. However, we have not yet developed a formal proof of either hypothesis. More generally, it might be possible to derive the error-handling heuristics from more-fundamental distributional assumptions.

Algorithmic Improvements for Question Design. There are a number of ways in which polyhedral question design might be improved. For example, the proposed algorithm reverts to random selection when the polyhedron becomes empty (around $q = 8$ when $p = 10$). We might explore algorithms that use relaxed constraints (OPT4) after the initial polyhedron becomes empty. Multistep look-ahead algorithms might yield further improvements.

Improved Hybrid Estimation. Analytic center estimation appears to do quite well when SE data are not available or when SE data are relatively noisy. However, when the SE data are relatively accurate, existing methods are better. In particular, for some domains a method that directly exploits the metric properties of the SEs seems to be best. We proposed ACSE as a natural way to mix AC estimation with SE constraints, but there might be other methods that make better use of the metric properties of the SEs. Table 4 also suggests that SE metric properties might be useful with hierarchical Bayes hybrids, such as the Ter Hofstede et al. (2002) algorithm.

Complexity Constraints. Evgeniou et al. (2002) demonstrate that "support vector machines" (SVM) can improve estimation by automatically balancing complexity of the partworth specification with fit. These researchers are now exploring a hybrid between analytic center and SVM estimation for stated-choice data—an exciting development that can

deal with nonlinearities in polyhedral specifications. We might also extend the algorithm in Appendix 1 to incorporate complexity considerations directly either by using relaxed constraints (thick hyperplanes as in Figure 2b) or the direct use of OPT4 in the definition of the hyperplanes.

Monte Carlo Experiments

Addressing the "Why." Our heuristics are designed to obtain better questions by reducing the feasible polyhedron as rapidly as possible. In an analogy to the theory of efficient questions, e.g., D-efficiency, this heuristic focuses the next question on that portion of the parameter space where we have the least information. Future Monte Carlo experiments might test this hypothesis or identify an alternative explanation. Such theory might lead to further improvements in the algorithm.

Learning and Wear-Out. The comparison of question order (polyhedral 1 versus polyhedral 2) suggests that the SE questions might increase the accuracy of the paired-comparison questions by acting as warm-up questions. This is an example of the more general issue that response errors might depend upon q. For example, exploratory simulations, using $\sigma_{pc} = \kappa_1 + \kappa_2 q$, suggest that learning (positive κ_2) causes mean absolute error (MAE) to decline more rapidly as the number of questions, q, increases. Wear-out (negative κ_2) yields a U-shaped function. Other simulations might explore further many behavioral issues affecting response accuracy (Tourangeau et al. 2000).

Interactions. Analytic center estimation is a by-product of polyhedral question design. Tables 1, 6, and 7 suggest that there might be interactions among question-design and estimation methods in terms of MAE or predictive ability. Although none of these interactions was significant in our data, interactions are worth further exploration. For example, while AC does well when the SE data are unavailable or relatively noisy, ACSE does not do as well for relatively accurate SE data.

Other Domains. Our simulations explore heterogeneity, response error, and the number of questions. We made a number of decisions such as the manner in which we specified heterogeneity (normally distributed with mean at the center of the range) and response error (normally distributed with no bias). Initial simulations suggested that the results were not sensitive to these assumptions, but more systematic exploration might identify interesting phenomena that we were not able to explore. Simulations for extremely large p or q might also yield new insight.

Other Criteria. We chose MAE as our evaluative criteria. Exploratory simulation suggested that root mean squared error (RMSE) appeared to be proportional to MAE. We might also explore other criteria such as predictive accuracy (analogous to Table 6), the dollar value of product features (Jedidi et al. 2003, Ofek and Srinivasan 2002), or the incremental explanatory power of each question and question type.

Empirical Tests in Other Categories

Known Applications. Polyhedral methods are beginning to diffuse. Sawtooth Software, Inc. now offers a polyhedral option to its ACA software, Harris Interactive, Inc. has begun initial testing, and National Family Opinion, Inc. is exploring feasibility. Sawtooth Software has completed an empirical test of internal validity using a poly/ACA question design algorithm (Orme and King 2002). In their data, on average, the ACA portion chose 63% of the paired-comparison questions. They observed no significant differences between the methods after $q = 30$. We have not been able to obtain for their application estimates of heterogeneity, PC response error, SE response error, or performance for low q.

Other Product Categories. We choose a category with separable features. In this category, the SE data were relatively accurate, and thus the ACA benchmark was not handicapped. This domain matched one of the $3 \times 2 \times 2$ classes of categories in Table 4, and the empirical data were consistent with the Monte Carlo data. We hypothesize that the Monte Carlo experiments are consistent with the 11 other classes of categories, but further empirical tests might explore this hypothesis further.

Acknowledgments

The authors gratefully acknowledge the contribution of Robert M. Freund, who proposed the use of the analytic center and approximating ellipsoids and gave us detailed advice on the application of these methods. This research was supported by the Sloan School of Management and the Center for Innovation in Product Development at MIT. The authors thank Limor Weisberg for creating the graphics used on the Web-based questionnaire, Rob Hardy, Adlar Kim, and Bryant Lin, who assisted in the development of the Web-based questionnaire, and Evan Davidson for documenting and improving the code. Bryan Orme and Richard Johnson at Sawtooth Software provided many useful comments on the empirical test and generously cooperated in providing information about Sawtooth Software's ACA algorithm and verification of our code. The study benefited from comments by numerous pretests 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. This paper has benefited from presentations at the CIPD Spring Research Review, the Epoch Foundation Workshop, the Marketing Science Conference in Wiesbaden Germany, the MIT ILP Symposium on Managing Corporate Innovation, the MIT Marketing Workshop, the MIT Operations Research Seminar Series, the MSI Young Scholars Conference, the New England Marketing Conference, and the Stanford Marketing Workshop. Open source code and data are available at mitsloan.mit.edu/vc and at mktsci.pubs.informs.org. Demos are available at mitsloan.mit.edu.

Appendix 1: Mathematics of Fast Polyhedral Adaptive Conjoint Estimation

Consider the case of p parameters and q questions where $q \le p$. Let u_j be the jth parameter of the respondent's partworth function and let $\vec{u}$ be the $p \times 1$ vector of parameters. Without loss of generality we assume binary features such that u_j is the high level of the jth feature and constrain their values between 0 and 100. For more levels we simply recode the $\vec{u}$ vector and impose constraints such as $u_m \le u_h$. We handle such inequality constraints by adding slack variables, $v_{hm} \ge 0$, such that $u_h = u_m + v_{hm}$. Let r be the number of externally imposed constraints, of which $r' \le r$ are inequality constraints.

Let $\vec{z}_{il}$ be the $1 \times p$ vector describing the left-hand profile in the ith paired-comparison question and let $\vec{z}_{ir}$ be the $1 \times p$ vector describing the right-hand profile. The elements of these vectors are binary indicators taking on the values 0 or 1. Let X be the $q \times p$ matrix of $\vec{x}_i = \vec{z}_{il} - \vec{z}_{ir}$ for $i = 1$ to q. Let a_i be the respondent's answer to the ith question and let $\vec{a}$ be the $q \times 1$ vector of answers for $i = 1$ to q. Then, if there were no errors, the respondent's answers imply $X\vec{u} = \vec{a}$. To handle additional constraints, we augment these equations such that X becomes a $(q + r) \times (p + r')$ matrix, $\vec{a}$ becomes a $(q + r) \times 1$ vector, and $\vec{u}$ becomes a $(p + r') \times 1$ vector. These augmented relationships form a polyhedron, $\mathsf{P} = \{\vec{u} \in \Re^{p+r'} \mid X\vec{u} = \vec{a},\ \vec{u} \ge 0\}$. We begin by assuming that P is nonempty, that X is full-rank, and that no j exists such that $u_j = 0$ for all $\vec{u}$ in P. We later indicate how to handle these cases.

Finding an Interior Point of the Polyhedron

To begin the algorithm we first find a feasible interior point of P by solving a linear program, (LP1) (Freund et al. 1985). Let $\vec{e}$ be a $(p + r') \times 1$ vector of 1s and let $\vec{0}$ be a $(p + r') \times 1$ vector of 0s; the y_js and θ are parameters of (LP1) and $\vec{y}$ is the $(p + r') \times 1$ vector of the y_js. (When clear in context, inequalities applied to vectors apply for each element.) (LP1) is given by

$$\begin{aligned} \max \quad & \sum_{j=1}^{p+r'} y_j, \\ \text{subject to} \quad & X\vec{u} = \theta\vec{a}, \qquad \theta \ge 1, \qquad \vec{u} \ge \vec{y} \ge \vec{0}, \qquad \vec{y} \le \vec{e}. \end{aligned} \tag{LP1}$$

If $(\vec{u}^*, \vec{y}^*, \theta^*)$ solves (LP1), then $\theta^{*-1}\vec{u}^*$ is an interior point of P whenever $\vec{y}^* > \vec{0}$. If there are some y_js equal to 0, then there are some js for which $u_j = 0$ for all $\vec{u} \in \mathsf{P}$. If LP1 is infeasible, then P is empty. We address these cases later in this appendix.

Finding the Analytic Center

The analytic center is the point in P that maximizes the geometric mean of the distances from the point to the faces of P. We find the analytic center by solving (OPT1).

$$\begin{aligned} \max \quad & \sum_{j=1}^{p+r'} \ln(u_j), \\ \text{subject to} \quad & X\vec{u} = \vec{a}, \qquad \vec{u} > \vec{0}. \end{aligned} \tag{OPT1}$$

Freund (1993) proves with projective methods that a form of Newton's method will converge rapidly for (OPT1). To implement Newton's method, we begin with the feasible point from (LP1) and improve it with a scalar, α, and a direction, $\vec{d}$, such that $\vec{u} + \alpha\vec{d}$ is close to the optimal solution of (OPT1). ($\vec{d}$ is a $(p + r') \times 1$ vector of d_js.) We then iterate subject to a stopping rule.

We first approximate the objective function with a quadratic expansion in the neighborhood of $\vec{u}$.

$$\sum_{j=1}^{p+r'} \ln(u_j + d_j) \approx \sum_{j=1}^{p+r'} \ln(u_j) + \sum_{j=1}^{p+r'} \left(\frac{d_j}{u_j} - \frac{d_j^2}{2u_j^2} \right). \tag{A1}$$

If we define U as a $(p + r') \times (p + r')$ diagonal matrix of the u_js, then the optimal direction solves (OPT2):

$$\begin{aligned} \max \quad & \vec{e}^T U^{-1}\vec{d} - (1/2)\vec{d}^T U^{-2}\vec{d}, \\ \text{subject to} \quad & X\vec{d} = \vec{0}. \end{aligned} \tag{OPT2}$$

Newton's method solves (OPT1) quickly by exploiting an analytic solution to (OPT2). To see this, consider first the Karush-Kuhn-Tucker (KKT) conditions for (OPT2). If $\vec{z}$ is a $(p + r') \times 1$ vector parameter of the KKT conditions that is unconstrained in sign, then the KKT conditions are written as

$$U^{-2}\vec{d} - U^{-1}\vec{e} = X^T\vec{z}, \tag{A2}$$

$$X\vec{d} = \vec{0}. \tag{A3}$$

Multiplying (A2) on the left by XU^2 gives $X\vec{d} - XU\vec{e} = XU^2X^T\vec{z}$. Applying (A3) to this equation gives: $-XU\vec{e} = XU^2X^T\vec{z}$. Because $U\vec{e} = \vec{u}$ and since $X\vec{u} = \vec{a}$, we have $-\vec{a} = XU^2X^T\vec{z}$. Because X is full rank and U is positive, we invert XU^2X^T to obtain $\vec{z} = -(XU^2X^T)^{-1}\vec{a}$. Now replace $\vec{z}$ in (A2) by this expression and multiply by U^2 to obtain $\vec{d} = \vec{u} - U^2X^T(XU^2X^T)^{-1}\vec{a}$.

According to Newton's method, the new estimate of the analytic center, $\vec{u}'$, is given by $\vec{u}' = \vec{u} + \alpha\vec{d} = U(\vec{e} + \alpha U^{-1}\vec{d})$. There are two cases for α. If $\| U^{-1}\vec{d} \| < 1/4$, then we use $\alpha = 1$ because $\vec{u}$ is already close to optimal and $\vec{e} + \alpha U^{-1}\vec{d} > \vec{0}$. Otherwise, we compute α with a line search.

Special Cases

If X is not full rank, XU^2X^T might not invert. We can either select questions such that X is full rank or we can make it so by removing redundant rows. Suppose that $\vec{x}_k$ is a row of X such that $\vec{x}_k^T = \sum_{i=1,\, i\neq k}^{q+r} \beta_i \vec{x}_i^T$. Then if $a_k = \sum_{i=1,\, i\neq k}^{q+r} \beta_i a_i$, we remove $\vec{x}_k$. If $a_k \neq \sum_{i=1,\, i\neq k}^{q+r} \beta_i a_i$, then P is empty and we employ (OPT4) described later in this appendix.

If in (LP1) we detect cases where some y_js $= 0$, then there are some js for which $u_j = 0$ for all $\vec{u} \in \mathsf{P}$. In the later case, we can still find the analytic center of the remaining polyhedron by removing those js and setting $u_j = 0$ for those indices. If P is empty we employ (OPT4).

Finding the Ellipsoid and Its Longest Axis

If $\vec{\tilde{u}}$ is the analytic center and $\overline{U}$ is the corresponding diagonal matrix, then Sonnevend (1985a, b) demonstrates that $\mathsf{E} \subseteq \mathsf{P} \subseteq \mathsf{E}_{p+r'}$, where $\mathsf{E} = \{\vec{u} \mid X\vec{u} = \vec{a},\ \sqrt{(\vec{u} - \vec{\tilde{u}})^T \overline{U}^{-2} (\vec{u} - \vec{\tilde{u}})} \leq 1\}$ and $\mathsf{E}_{p+r'}$ is constructed proportional to E by replacing 1 with $(p + r')$. Because we are interested only in the direction of the longest axis of the ellipsoids we can work with the simpler of the proportional ellipsoids, E. Let $\vec{g} = \vec{u} - \vec{\tilde{u}}$, then the longest axis will be a solution to (OPT3).

$$\begin{aligned} \max \quad & \vec{g}^T\vec{g}, \\ \text{subject to} \quad & \vec{g}^T\overline{U}^{-2}\vec{g} \leq 1, \qquad X\vec{g} = \vec{0}. \end{aligned} \tag{OPT3}$$

(OPT3) has an easy-to-compute solution based on the eigenstructure of a matrix. To see this we begin with the KKT conditions (where ϕ and γ are parameters of the conditions)

$$\vec{g} = \phi\overline{U}^{-2}\vec{g} + X^T\vec{\gamma}, \tag{A4}$$

$$\phi(\vec{g}^T\overline{U}^{-2}\vec{g} - 1) = 0, \tag{A5}$$

$$\vec{g}^T\overline{U}^{-2}\vec{g} \leq 1, \qquad X\vec{g} = \vec{0}, \qquad \phi \geq 0. \tag{A6}$$

It is clear that $\vec{g}^T\overline{U}^{-2}\vec{g} = 1$ at optimal, else we could multiply $\vec{g}$ by a scalar greater than 1 and still have $\vec{g}$ feasible. It is likewise clear that ϕ is strictly positive, else we obtain a contradiction by left-multiplying (A4) by $\vec{g}^T$ and using $X\vec{g} = \vec{0}$ to obtain $\vec{g}^T\vec{g} = 0$, which contradicts $\vec{g}^T\overline{U}^{-2}\vec{g} = 1$. Thus, the solution to (OPT3) must satisfy $\vec{g} = \phi\overline{U}^{-2}\vec{g} + X^T\vec{\gamma}$, $\vec{g}^T\overline{U}^{-2}\vec{g} = 1$, $X\vec{g} = \vec{0}$, and $\phi > 0$. We rewrite (A4)–(A6) by letting I be the identify matrix and defining $\eta = 1/\phi$ and $\vec{\omega} = -\vec{\gamma}/\phi$.

$$(\overline{U}^{-2} - \eta I)\vec{g} = X^T\vec{\omega}, \tag{A7}$$

$$\vec{g}^T\overline{U}^{-2}\vec{g} = 1, \tag{A8}$$

$$X\vec{g} = \vec{0}, \qquad \phi > 0. \tag{A9}$$

We left-multiply (A7) by X and use (A9) to obtain $X\overline{U}^{-2}\vec{g} = XX^T\vec{\omega}$. Because X is full rank, XX^T is invertible, and we obtain $\vec{\omega} = (XX^T)^{-1}X\overline{U}^{-2}\vec{g}$, which we substitute into (A7) to obtain $(\overline{U}^{-2} - X^T(XX^T)^{-1}X\overline{U}^{-2})\vec{g} = \eta\vec{g}$. Thus, the solution to (OPT3) must be an eigenvector of the matrix, $M \equiv (\overline{U}^{-2} - X^T(XX^T)^{-1}X\overline{U}^{-2})$. To find out which eigenvector, we left-multiply (A7) by $\vec{g}^T$ and use (A8) and (A9) to obtain $\eta\vec{g}^T\vec{g} = 1$, or $\vec{g}^T\vec{g} = 1/\eta$ where $\eta > 0$. Thus, to solve (OPT3) we maximize $1/\eta$ by selecting the smallest positive eigenvalue of M. The direction of the longest axis is then given by the associated eigenvector of M. We then choose the next question such that $\vec{x}_{q+1}$ is most nearly collinear to this eigenvector subject any constraints imposed by the questionnaire design. (For example, in our simulation we require that the elements of $\vec{x}_{q+1}$ be -1, 0, or 1.) The answer to $\vec{x}_{q+1}$ defines a hyperplane orthogonal to $\vec{x}_{q+1}$.

We need only establish that the eigenvalues of M are real. To do this we recognize that $M = P\overline{U}^{-2}$ where $P = (I - X^T(XX^T)^{-1}X)$ is symmetric, i.e., $P = P^T$. Then if η is an eigenvalue of M, $\det[P\overline{U}^{-2} - \eta I) = 0$, which implies that $\det[\overline{U}(\overline{U}^{-1}P\overline{U}^{-1} - \eta I)\overline{U}^{-1}] = 0$. This implies that η is an eigenvalue of $\overline{U}^{-1}P\overline{U}^{-1}$, which is symmetric. Thus, η is real (Hadley 1961, p. 240).

Adjusting the Polyhedron so That It Is Nonempty

P will remain nonempty as long as respondents' answers are consistent. However, in any real situation there is likely to be $q < p$ such that P is empty. To continue the polyhedral algorithm, we adjust P so that it is nonempty. We do this by replacing the equality constraint, $X\vec{u} = \vec{a}$, with two inequality constraints, $X\vec{u} \leq \vec{a} + \vec{\delta}$ and $X\vec{u} \geq \vec{a} - \vec{\delta}$, where $\vec{\delta}$ is a $q \times 1$ vector of errors, δ_i, defined only for the question-answer imposed constraints. We solve the following optimization problem. Our current implementation uses the ∞-norm, where we minimize the maximum δ_i, but other norms are possible. The advantage of using the ∞-norm is that (OPT4) is solvable as a linear program.

$$\begin{aligned} \min \quad & \|\delta\| \\ \text{subject to} \quad & X\vec{u} \leq \vec{a} + \vec{\delta}, \quad X\vec{u} \geq \vec{a} - \vec{\delta}, \quad \vec{u} \geq \vec{0}. \end{aligned} \tag{OPT4}$$

At some point such that $q > p$, extant algorithms will outperform (OPT4) and we can switch to those algorithms. Alternatively, a researcher might choose to switch to constrained regression (norm-2) or mean-absolute error (norm-1) when $q > p$. Other options include replacing some, but not all, of the equality constraints with inequality constraints. We leave these extensions to future research.

Appendix 2: Internal Validity Tests for Laptop-Computer Bags

Table A2.1 Correlation with Actual Response

	After 8 Questions		After 16 Questions	
Methods Without SE Data	Fixed Questions	Polyhedral 1 Questions	Fixed Questions	Polyhedral 1 Questions
Analytic center (AC)	0.65	0.69	0.80	0.79
Hierarchical Bayes (HB)	0.70	0.67	0.76	0.72
Sample size	88	87	88	87
Methods that Use SE Data	ACA Questions	Polyhedral 2 Questions	ACA Questions	Polyhedral 2 Questions
WHSE estimation	0.77	0.81	0.81	0.84
ACSE estimation	0.74	0.78	0.77	0.84
HBSE estimation	0.76	0.80	0.78	0.82
Sample size	80	71	80	71

Note. The missing observations reflect respondents who gave the same response for all four holdout questions (in which case the correlations were undefined).

Table A2.2 Conclusions from the Multivariate Analysis

	Without SE Questions	With SE Questions
Comparison of Estimation Methods		
Fixed questions	HB > AC	
Polyhedral 1 questions	AC >>> HB	
ACA questions		WHSE > HBSE >>> ACSE
Polyhedral 2 questions		WHSE > HBSE > ACSE
Comparison of Question-Design Methods		
AC estimation	Polyhedral 1 > Fixed	
HB estimation	Fixed > Polyhedral 1	
WHSE estimation		Polyhedral 2 >>> ACA
ACSE estimation		Polyhedral 2 >>> ACA
HBSE estimation		Polyhedral 2 >>> ACA

References

Allenby, Greg M., Peter E. Rossi. 1999. Marketing models of consumer heterogeneity. *J. Econometrics* **89**(March/April) 57–78.

Andrews, Rick L., Andrew Ainslie, Imran S. Currim. 2002. An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity. *J. Marketing Res.* **39**(November) 479–487.

Arora, Neeraj, Greg M. Allenby, James L. Ginter. 1998. A hierarchical Bayes model of primary and secondary demand. *Marketing Sci.* **17**(1) 29–44.

——, Joel Huber. 2001. Improving parameter estimates and model prediction by aggregate customization in choice experiments. *J. Consumer Res.* **28**(September) 273–283.

Bucklin, Randolph E., V. Srinivasan. 1991. Determining interbrand substitutability through survey measurement of consumer preference structures. *J. Marketing Res.* **28**(February) 58–71.

Currim, Imran S. 1981. Using segmentation approaches for better prediction and understanding from consumer mode choice models. *J. Marketing Res.* **18**(August) 301–309.

Dahan, Ely, John R. Hauser. 2002. The virtual customer. *J. Product Innovation Management* **19**(5) 332–354.

Dawes, Robin M., Bernard Corrigan. 1974. Linear models in decision making. *Psych. Bull.* **81**(March) 95–106.

Einhorn, Hillel J. 1971. Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organ. Behavior Human Performance* **6** 1–27.

Elrod, Terry, Jordan Louviere, Krishnakumar S. Davey. 1992. An empirical comparison of ratings-based and choice-based conjoint models. *J. Marketing Res.* **29**(3)(August) 368–377.

Evgeniou, Theodoros, Constantinos Boussios, Giorgos Zacharia. 2002. Generalized robust conjoint estimation. Working paper, INSEAD, Fontainebleau, France.

Freund, Robert. 1993. Projective transformations for interior-point algorithms, and a superlinearly convergent algorithm for the W-center problem. *Math. Programming* **58** 385–414.

——, R. Roundy, M. J. Todd. 1985. Identifying the set of always-active constraints in a system of linear inequalities by a single linear program. WP 1674-85, MIT Sloan School of Management, Cambridge, MA.

Green, Paul E. 1984. Hybrid models for conjoint analysis: An expository review. *J. Marketing Res.* **21**(2) 155–169.

——, V. Srinivasan. 1978. Conjoint analysis in consumer research: Issues and outlook. *J. Consumer Res.* **5**(2) 103–123.

——, ——. 1990. Conjoint analysis in marketing: New developments with implications for research and practice. *J. Marketing* **54**(4) 3–19.

——, Kristiaan Helsen, Bruce Shandler. 1988. Conjoint internal validity under alternative profile presentations. *J. Consumer Res.* **15**(December) 392–397.

——, Stephen M. Goldberg, Mila Montemayor. 1981. A hybrid utility estimation model for conjoint analysis. *J. Marketing* 33–41.

——, Abba Krieger, Manoj K. Agarwal. 1991. Adaptive conjoint analysis: Some caveats and suggestions. *J. Marketing Res.* **23**(2) 215–222.

——, ——, Jerry Wind. 2001. Thirty years of conjoint analysis: Reflections and prospects. *Interfaces* **21**(3, Part 2 of 2) (May–June) S56–S73.

——, ——, ——. 2002. Buyer choice simulators, optimizers, and dynamic models. *Advances in Marketing Research: Progress and Prospects*. Wharton Press, Philadelphia, PA.

Griffin, Abbie, John R. Hauser. 1993. The voice of the customer. *Marketing Sci.* **12**(1) 1–27.

Gritzmann, P., V. Klee. 1993. Computational complexity of inner and outer J-radii of polytopes in finite-dimensional normed spaces. *Math. Programming* **59**(2) 163–213.

Hadley, G. 1961. *Linear Algebra.* Addison-Wesley Publishing Company, Inc., Reading, MA.

Hauser, John R., Steven P. Gaskin. 1984. Application of the "Defender" consumer model. *Marketing Sci.* **3**(4) 327–351.

——, Vithala Rao. 2003. Conjoint analysis, related modeling, and applications. Jerry Wind, ed. *Advances in Marketing Research: Progress and Prospects*. Forthcoming.

——, Steven M. Shugan. 1980. Intensity measures of consumer preference. *Oper. Res.* **28**(2) 278–320.

Hoaglin, David C., Frederick Mosteller, John W. Tukey. 1983. *Understanding Robust and Exploratory Data Analysis*. John Wiley & Sons, Inc., New York.

Huber, Joel. 1975. Predicting preferences on experimental bundles of attributes: A comparison of models. *J. Marketing Res.* **12**(August) 290–297.

——, Klaus Zwerina. 1996. The importance of utility balance in efficient choice designs. *J. Marketing Res.* **33**(August) 307–317.

——, Dick R. Wittink, John A. Fiedler, Richard Miller. 1993. The effectiveness of alternative preference elicitation procedures in predicting choice. *J. Marketing Res.* **30**(1) 105–114.

Jedidi, Kamel, Puneet Manchanda, Sharan Jagpal. 2003. Measuring heterogeneous reservation prices for product bundles. *Marketing Sci.* **22**(1). Forthcoming.

Johnson, Eric J., Robert J. Meyer, Sanjoy Ghose. 1989. When choice models fail: Compensatory models in negatively correlated environments. *J. Marketing Res.* **26**(3) 255–270.

Johnson, Richard. 1987. Accuracy of utility estimation in ACA. Working paper, Sawtooth Software, Sequim, WA.

——. 1991. Comment on "Adaptive conjoint analysis: Some caveats and suggestions." *J. Marketing Res.* **28**(May) 223–225.

——. 1999. The joys and sorrows of implementing HB methods for conjoint analysis. Working paper, Sawtooth Software, Sequim, WA.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl, T. C. Lee. 1985. *The Theory and Practice of Econometrics*. John Wiley and Sons, New York, 571.

Karmarkar, N. 1984. A new polynomial time algorithm for linear programming. *Combinatorica* **4** 373–395.

Kuhfeld, Warren F., Randall D. Tobias, Mark Garratt. 1994. Efficient experimental design with marketing research applications. *J. Marketing Res.* **31**(4) 545–557.

Leigh, Thomas W., David B. MacKay, John O. Summers. 1984. Reliability and validity of conjoint analysis and self-explicated weights: A comparison. *J. Marketing Res.* **21**(4) 456–462.

Lenk, Peter J., Wayne S. DeSarbo, Paul E. Green, Martin R. Young. 1996. Hierarchical Bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Sci.* **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. *J. Marketing Res.* **38**(2) 183–196.

Louviere, Jordan J., David A. Hensher, Joffre D. Swait. 2000. *Stated Choice Methods: Analysis and Application*. Cambridge University Press, New York.

McFadden, Daniel. 2000. Disaggregate behavioral travel demand's RUM side: A thirty-year retrospective. Working paper, University of California, Berkeley, CA.

Moore, William L. 2003. A cross-validity comparison of conjoint analysis and choice models. Working paper, University of Utah, Salt Lake City, UT.

——, Richard J. Semenik. 1988. Measuring preferences with hybrid conjoint analysis: The impact of a different number of attributes in the master design. *J. Bus. Res.* 261–274.

Nesterov, Y., A. Nemirovskii. 1994. Interior-point polynomial algorithms in convex programming. *SIAM*, Philadelphia, PA.

Ofek, Elie, V. Srinivasan. 2002. How much does the market value an improvement in a product attribute? *Marketing Sci.* **21**(4) 98–411.

Orme, Bryan. 1999. ACA, CBC, or both?: Effective strategies for conjoint research. Working paper, Sawtooth Software, Sequim, WA.

——, W. Christopher King. 2002. Improving ACA algorithms: Challenging a twenty-year-old approach. *Pres. Adv. Res. Tech. Conf.*, Vail, CO, June 3.

Page, Albert L., Harold F. Rosenbaum. 1989. Redesigning product lines with conjoint analysis: How Sunbeam does it. *J. Product Innovation Management* **4** 120–137.

Reibstein, David, John E. G. Bateson, William Boulding. 1988. Conjoint analysis reliability: Empirical findings. *Marketing Sci.* **7**(3) 271–286.

Robinson, P. J. 1980. Applications of conjoint analysis to pricing problems. David B. Montgomery, Dick Wittink, eds. *Market Measurement and Analysis: Proc. 1979 ORSA/TIMS Conf. Marketing*. Marketing Science Institute, Cambridge, MA, 183–205.

Sandor, Zsolt, Michel Wedel. 2001. Designing conjoint choice experiments using managers' prior beliefs. *J. Marketing Res.* **38**(4) 430–444.

——, ——. 2002. Profile construction in experimental choice designs for mixed logit models. *Marketing Sci.* **21**(4) 398–411.

Sawtooth Software, Inc. 1996. ACA system: Adaptive conjoint analysis. *ACA Manual*. Sequim, WA.

——. 2001. The ACA/hierarchical Bayes technical paper. Sawtooth Software, Inc., Sequim, WA.

——. 2002. ACA 5.0 Technical paper. Sawtooth Software, Inc., Sequim, WA.

Sonnevend, G. 1985a. An "analytic" center for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. *Proc. 12th IFIP Conf. System Modeling and Optim.* Budapest, Hungary.

——. 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, Eötvös University, Budapest, Hungary.

Srinivasan, V. 1988. A conjunctive-compensatory approach to the self-explication of multiattributed preferences. *Decision Sci.* **19**(2) 295–305.

——, Chan Su Park. 1997. Surprising robustness of the self-explicated approach to customer preference structure measurement. *J. Marketing Res.* **34**(May) 286–291.

——, Allan D. Shocker. 1973. Linear programming techniques for multidimensional analysis of preferences. *Psychometrika* **38**(3) 337–369.

Ter Hofstede, Frenkel, Youngchan Kim, Michel Wedel. 2002. Bayesian prediction in hybrid conjoint analysis. *J. Marketing Res.* **39**(May) 253–261.

Tourangeau, Roger, Lance J. Rips, Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge University Press, New York, 197–229.

Toubia, Olivier, Duncan Simester, John R. Hauser. 2003. Polyhedral methods for adaptive choice-based conjoint analysis. *J. Marketing Res.* Forthcoming.

Tukey, John W. 1960. A survey of sampling from contaminated distributions. I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, H. Mann, eds. *Contributions to Probability and Statistics*. Stanford University Press, Stanford, CA, 448–485.

Urban, Glen L., Gerald M. Katz. 1983. Pre-test market models: Validation and managerial implications. *J. Marketing Res.* **20**(August) 221–234.

Vaidja, P. 1989. A locally well-behaved potential function and a simple Newton-type method for finding the center of a polytope. N. Megiddo, ed. *Progress in Mathematical Programming: Interior Points and Related Methods*. Springer, New York, 79–90.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48** 817–838.

Wittink, Dick R., David B. Montgomery. 1979. Predictive validity of trade-off analysis for alternative segmentation schemes. Neil Beckwith, ed. *Educators' Conf. Proc. 1979 AMA*. American Marketing Association, Chicago, IL.

——, Philippe Cattin. 1981. Alternative estimation methods for conjoint analysis: A Monte Carlo study. *J. Marketing Res.* **18**(February) 101–106.

Wright, Peter, Mary Ann Kriewall. 1980. State-of-mind effects on accuracy with which utility functions predict marketplace utility. *J. Marketing Res.* **17**(August) 277–293.

Yang, Sha, Greg M. Allenby, Geraldine Fennell. 2002. Modeling variation in brand preference: The roles of objective environment and motivating conditions. *Marketing Sci.* **21**(1) 14–31.