MARKETING SCIENCE
Vol. 26, No. 5, September–October 2007, pp. 596–610
ISSN 0732-2399 | EISSN 1526-548X | 07 | 2605 | 0596

informs

DOI 10.1287/mksc.1060.0257

Probabilistic Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis: Theory and Application

Olivier Toubia
Columbia Business School, Columbia University, 522 Uris Hall, 3022 Broadway, New York, New York 10027, ot2107@columbia.edu

John Hauser
MIT Sloan School of Management, Massachusetts Institute of Technology, 40-179, 1 Amherst Street, Cambridge, Massachusetts 02142, jhauser@mit.edu

Rosanna Garcia
College of Business Administration, Northeastern University, 202 HA, Boston, Massachusetts 02115, r.garcia@neu.edu

Polyhedral methods for choice-based conjoint analysis provide a means to adapt choice-based questions at the individual-respondent level and provide an alternative means to estimate partworths when there are relatively few questions per respondent, as in a Web-based questionnaire. However, these methods are deterministic and are susceptible to the propagation of response errors. They also assume, implicitly, a uniform prior on the partworths. In this paper we provide a probabilistic interpretation of polyhedral methods and propose improvements that incorporate response error and/or informative priors into individual-level question selection and estimation.

Monte Carlo simulations suggest that response-error modeling and informative priors improve polyhedral question-selection methods in the domains where they were previously weak. A field experiment with over 2,200 leading-edge wine consumers in the United States, Australia, and New Zealand suggests that the new question-selection methods show promise relative to existing methods.

Key words: conjoint analysis; choice models; estimation and other statistical techniques; international marketing; marketing research; new-product research; product development; Bayesian methods
History: This paper was received May 26, 2006, and was with the authors 1 month for 1 revision; processed by Eric Bradlow.

1. Introduction

Toubia et al. (2003) demonstrated that polyhedral methods for adaptively selecting questions in metric conjoint analysis could improve accuracy when partworths are either homogeneous or heterogeneous, and could do so whether response errors are large or small. Toubia et al. (THS 2004) extended polyhedral methods to choice-based questions, but with mixed success. Polyhedral choice-based questions improved accuracy when response errors were low, but not when they were high. Furthermore, although polyhedral methods for metric paired-comparison questions predict well for empirical data, there have been no empirical validity tests for choice-based polyhedral methods despite the growing interest among practitioners for adaptive choice-based methods.

In this paper we propose and test a generalization of THS that takes response error into account for choice-based questions and has the potential to improve accuracy in high response-error domains. We do so by recasting the polyhedral heuristic into a Bayesian framework. This framework also includes prior information in a natural, conjugate manner. After verifying the methods with simulations, we undertake a large-scale, multicountry study in which each respondent completes two separate conjoint tasks. This design enables us to compare question selection with a within-respondent design that implies greater statistical power to distinguish methods. We compare methods on the ability to predict actual choices. We examine whether the methods lead to different managerial implications by comparing forecasts of willingness to pay as well as the optimal product lines implied by each method.

This paper is organized as follows. Section 2 briefly reviews the published choice-based polyhedral methods and discusses two key limitations. Sections 3 and 4 propose solutions to these limitations. Sections 5 examines the methods with Monte Carlo simulations. Section 6 describes the methodological results of the field experiment. Section 7 concludes and offers directions for future research.

2. Review and Critique of Polyhedral Choice-Based Methods

Choice-based polyhedral question selection selects each choice question to learn as much as possible about a respondent's preferences. The conceptual idea is to recognize that the set of choice questions and their corresponding answers define a polyhedron, i.e., a set of "feasible" partworth vectors that perfectly fit previous observations. Each choice narrows the range of feasible partworths making the range smaller and smaller until it converges toward a single partworth vector. The method works well when the respondent makes no errors, but can be highly sensitive to errors, particularly in the early choices. We now provide a brief technical review to establish both notation and conceptual reasoning for the generalizations.

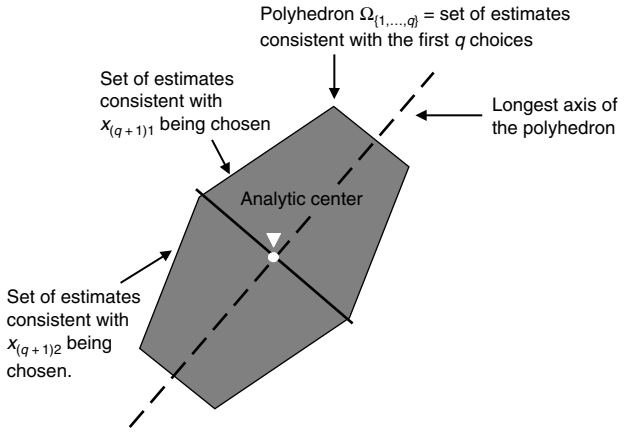
Answers to Choice-Based Questions Interpreted as Constraints on the Partworths

Without loss of generality, we use binary vectors in the theoretical development to simplify notation and exposition. Multilevel features are used in both the simulations and the application. Let x_{qjf} indicate that the jth alternative in the qth choice set contains the fth feature, and let $\vec{x}_{qj}$ be the binary row vector describing the jth alternative in the qth choice set. Define $\vec{x}_{qk}$ similarly for the kth profile. Let $\vec{u}$ be the l-dimensional vector of partworths for a given respondent. Let ε_{qj} and ε_{qk} be error terms such that the respondent's utility for profile j in choice set q is $\vec{x}_{qj}\vec{u} + \varepsilon_{qj}$. The utility-maximizing respondent will choose profile j^* over profile k if and only if $(\vec{x}_{qj^*} - \vec{x}_{qk})\vec{u} + (\varepsilon_{qj^*} - \varepsilon_{qk}) \geq 0$. Each choice among J alternatives implies $J-1$ such inequality constraints, indicating that the utility of the chosen profile is higher than that of the other $J-1$ alternatives in the choice set. Let $X_{\{1,\ldots,q\}}$ be the matrix of the $(\vec{x}_{qj^*} - \vec{x}_{qk})$s for all $J-1$ inequality constraints stacked for the first q questions. Note that the respondent's q choices are coded in $X_{\{1,\ldots,q\}}$ by the selection of j^* for each question. Let $\vec{\varepsilon}$ be the corresponding vector of error differences and, without loss of generality, scale all partworths to be nonnegative and normalize the partworths so that they sum to 100.[1] Then, if $\vec{e}$ is a vector of 1s and $\vec{0}$ is a vector of 0s (of length l), the answers to the choice-based questions imply the following constraints on the respondent's partworths:

$$\text{(P1)} \quad X_{\{1,\ldots,q\}}\vec{u} + \vec{\varepsilon} \geq \vec{0} \quad \vec{u} \geq \vec{0} \quad \vec{e}'\vec{u} = 100.$$

[1] Nonnegativity assumes that we know a priori which level of the partworth vector is preferred. This simplifies notation. We address empirical issues in later sections. The selection of 100 is arbitrary and implicitly rescales the error vector, $\vec{\varepsilon}$.

Figure 1 Deterministic Polyhedral Question Selection



Question Selection

For any given vector $\vec{\varepsilon}$, the set of vectors $\vec{u}$ satisfying the constraints in (P1) is a mathematical object called a polyhedron. THS select questions such that the polyhedron corresponding to $\vec{\varepsilon} = \vec{0}$ never becomes empty, and effectively assume that $\vec{\varepsilon} = \vec{0}$. Let $\Omega_{\{1,\ldots,q\}}$ be the polyhedron obtained after q questions. The $q+1$st question imposes new constraints on the partworths, giving rise to a new polyhedron $\Omega_{\{1,\ldots,q+1\}}$ that is a subset of the previous polyhedron $\Omega_{\{1,\ldots,q\}}$. For a linear compensatory utility model, each point in $\Omega_{\{1,\ldots,q\}}$ is consistent with the respondent choosing one and only one of the alternatives in choice set $(q+1)$ (except for a set of points of measure 0 for which the respondent is indifferent between at least two profiles). Hence, the $q+1$st question divides $\Omega_{\{1,\ldots,q\}}$ into J collectively exhaustive (smaller) polyhedra that are of roughly equal size. The region corresponding to the respondent's choice becomes the starting polyhedron for the next question. See Figure 1 for a choice set of two alternatives. If there were no response errors, the sequence of polyhedra would shrink toward the respondent's true partworth vector.

Question selection (choice set selection) obeys two principles: (1) choice balance and (2) postchoice symmetry. Choice balance minimizes the expected size of $\Omega_{\{1,\ldots,q+1\}}$ and is implemented by ensuring that a respondent who uses the working estimate of the partworths, $\hat{\vec{u}}_q$, would be approximately indifferent between all the alternatives in the choice set. Choice balance is common in the literature and, for choice questions, typically increases the efficiency of the questions (Arora and Huber 2001, Hauser and Toubia 2005, Huber and Zwerina 1996, Kanninen 2002).[2] Postchoice symmetry minimizes the maximum

[2] The first-order conditions for logit-based choice-based questions indicate that the information matrix is maximized for questions that are close to, but not perfectly, choice balanced. See appendix to Hauser and Toubia (2005).

uncertainty on any combination of partworths, and is implemented by constructing choice sets that divide the polyhedron $\Omega_{\{1,\ldots,q\}}$ perpendicularly to its longest axes.

Estimation

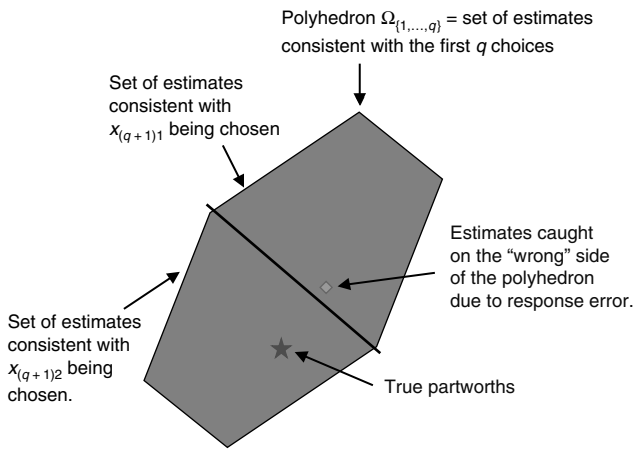
Because choice questions are chosen such that the polyhedron $\Omega_{\{1,\ldots,q\}}$ never becomes empty, all points in $\Omega_{\{1,\ldots,q\}}$ are consistent with all of the respondent's choices. Thus, THS use the analytic center of $\Omega_{\{1,\ldots,q\}}$, $\hat{\vec{u}}_q$, as the working estimate of $\vec{u}$ after q questions.

Critique

Choice-based polyhedral question selection and estimation are promising. Empirically, choice balance is achieved and the polyhedra shrink rapidly (although there is no published data on the ability to predict actual choices). Compared to randomly generated questions, orthogonal designs, and aggregate customization (Arora and Huber 2001, Huber and Zwerina 1996), deterministic choice-based polyhedral questions improve accuracy when response error is low, but not when response error is high.

The poor performance for high response errors is likely due to response-error propagation, as illustrated in Figure 2. In this example, the respondent's true partworth values are as indicated by a star (★). With no response error, the respondent would choose Profile 2, corresponding to the lower polyhedron, and the set of feasible partworths (new polyhedron) would converge toward the true value. However, with response errors the respondent might choose Profile 1, corresponding to the upper polyhedron. Once such a choice is made, the partworths can never converge to the true value. The closest estimate would be on the border, as indicated by the small diamond (♦). Moreover, without a formal probabilistic structure, there is no easy way to incorporate prior information on the likely distribution of partworths. We next address both response error and prior information with a Bayesian interpretation of choice-based polyhedral methods.

Figure 2 Illustration of Response Error in Deterministic Polyhedral Question Selection



3. Bayesian Interpretation for Choice-Based Polyhedral Methods

We can interpret the analytic center as a working estimate if we assume a prior distribution on the partworth vector $\vec{u}$ that is uniformly distributed on the initial polyhedron, $\Omega_0 = \{\vec{u}\colon\ \vec{u} \geq \vec{0}, \vec{e}'\vec{u} = 100\}$. Denote this distribution as P_{Ω_0}: defined by $P_{\Omega_0}(\vec{u}) = 0$ if $\vec{u} \notin \Omega_0$; $P_{\Omega_0}(\vec{u}) = \|\Omega_0\|^{-1}$ if $\vec{u} \in \Omega_0$. ($\|\Omega_0\|$ is the measure of the set Ω_0.) Denote the data provided by the respondent through the first q choices with $D_{\{1,\ldots,q\}}$. ($D_{\{1,\ldots,q\}}$ is encoded in $X_{\{1,\ldots,q\}}$.) The deterministic algorithm implicitly assumes a likelihood function of the form: $P(D_{\{1,\ldots,q\}} \mid \vec{u}) = 1$ if $\vec{u} \in \Omega_{\{1,\ldots,q\}}$ and $P(D_{\{1,\ldots,q\}} \mid \vec{u}) = 0$ if $\vec{u} \notin \Omega_{\{1,\ldots,q\}}$. Applying Bayes rule, $P(\vec{u} \mid D_{\{1,\ldots,q\}}) \propto P(D_{\{1,\ldots,q\}} \mid \vec{u}) P_{\Omega_0}(\vec{u}) \propto P_{\Omega_{\{1,\ldots,q\}}}(\vec{u})$. In other words, the posterior distribution is the uniform distribution with support $\Omega_{\{1,\ldots,q\}}$.

Once the method is viewed in a Bayesian framework, the two implicit assumptions of the absence of response error and uniform priors may easily be relaxed by generalizing, respectively, the likelihood function and the prior.

4. Probabilistic Polyhedral Methods

Generalizing the Likelihood Function

In the deterministic algorithm $\vec{\varepsilon} = \vec{0}$, and the respondent chooses the profile with the highest deterministic utility with probability 1. All posterior distributions are uniform distributions supported by polyhedra. We generalize the algorithm by considering distributions supported by mixtures of polyhedra. As the number of polyhedra in the mixtures grows, we can approximate any distribution, but we must balance this capability with the realization that as the number of polyhedra grows, the computational time grows exponentially. To balance these effects, we choose a simple likelihood function that captures the essence of response error. We use simulations to examine whether this is a sufficient approximation.

To obtain a structure in which the prior and posterior distributions are conjugate, we assume that the noise $\vec{\varepsilon}$ is distributed such that the respondent chooses the profile with the highest deterministic utility with probability, α', and chooses the $J-1$ other profiles with probability $(1-\alpha')/(J-1)$. The advantages of this assumption are that it provides a feasible algorithm and nests the deterministic algorithm as the special case when $\alpha' = 1$. While we believe this assumption is a reasonable, first-order robust

assumption, it may not hold exactly in real or synthetic data. To test the robustness of this assumption, we generate data in our simulations that use a traditional logistic function and, hence, violate this assumption to some degree.

In general, α' is unknown and can be assumed to vary across respondents and, potentially, across choice sets within a respondent (e.g., α' may be higher or lower if the profiles in the choice set are closer in utility). We might include priors for α' and do a full Bayesian updating such that $P(\vec{u}, \{\alpha'_i\} \mid D_{\{1,\ldots,q\}}, \{x_{i1}, x_{i2}, \ldots, x_{iJ}\}) \propto [\prod_i P(D_i \mid \vec{u}, \alpha'_i) P(\alpha'_i \mid \vec{u}, x_{i1}, x_{i2}, \ldots, x_{iJ})] P(\vec{u})$.

To avoid complexity, for a first test of the algorithm, we model α' as homogeneous and constant. Fortunately, sensitivity analyses suggest that predictive ability is not sensitive to the choice of α' within a wide range that is consistent with the α's estimated for our simulations and empirical test. See Appendices A1 and A2 for details on estimation and sensitivity. We leave to future research the investigation of alternative ways to specify and estimate α'. For our empirical tests, we use pretest data to select a point estimate of α'. Pretest selection follows the tradition of aggregate customization (Arora and Huber 2001, Huber and Zwerina 1996).

Generalizing the Prior Distribution

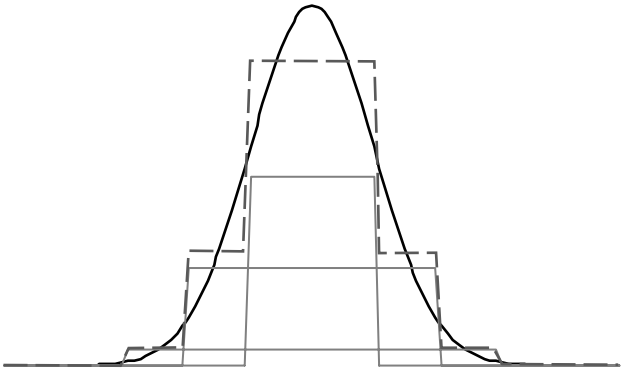
We nest THS's implicit prior distribution within a mixture of uniform distributions supported by polyhedra: $\sum_{m=1}^{M} \omega_m P_{\Psi_m}(\vec{u})$ where M is any positive integer, $\{\omega_1, \ldots, \omega_M\}$ is a set of positive weights such that $\sum_{m=1}^{M} \omega_m = 1$, and $\{\Psi_1, \ldots, \Psi_M\}$ is a set of polyhedra.

In this paper we apply and test two special cases of nonuniform priors. The first special case approximates traditional normal priors. Figure 3 illustrates the approximation conceptually. (In one dimension, a polyhedron is a line segment.) The uniform distributions are indicated with solid lines; the approximation with a dotted line. Appendix A3 provides a procedure for choosing weights for the polyhedra providing support for the distribution.

The second special case, denoted "population priors," selects a mixture of polyhedra such that the median of the prior importance of each feature is equal to the median (across respondents) of its importance. The polyhedra are defined by inequalities implied by the median importances of the features. If F is the number of features, this prior uses a mixture of 2^F polyhedra.[3] Details on the definitions of the polyhedra and the computation of the weights are given in Appendix 4.

[3] The importance can be defined as the difference between the maximum and minimum partworths for a feature or the average absolute magnitudes of the partworths. Importances have both conceptual and computational advantages relative to imposing median constraints on the partworths directly.

Figure 3 Approximating a Normal Prior with a Mixture of Polyhedra



Conjugate Posterior Distributions

An important feature of our generalization is that the class of likelihood functions and the class of priors presented above are conjugate; that is, the posterior distributions remain within the set of mixtures of uniform distributions supported by polyhedra.

In order to show this result, we begin with a prior distribution, $P_{\Omega_0}(\vec{u})$, supported by a (single) polyhedron, Ω_0. Let $\Omega_1 = \{\vec{u}: X_1\vec{u} \geq \vec{0}, \vec{u} \geq \vec{0}, \vec{e}'\vec{u} = 100\}$ be the polyhedron defined by the answer to the first question and D_1 be the data provided by this question. Let $\Omega_0 - \Omega_1$ be the set in which all points in Ω_1 are removed from Ω_0. Applying Bayes rule:

$$p(\vec{u} \mid D_1) \propto p(D_1 \mid \vec{u}) P_{\Omega_0}(\vec{u}) = \begin{cases} \alpha' \|\Omega_0\|^{-1} & \text{if } \vec{u} \in \Omega_1 \\ \left(\dfrac{1-\alpha'}{J-1}\right) \|\Omega_0\|^{-1} & \text{if } \vec{u} \in \Omega_0 - \Omega_1. \end{cases} \quad (1)$$

The posterior is proportional to a piecewise-constant function that takes the values of zero at all points outside Ω_0, $\alpha' \|\Omega_0\|^{-1}$ at all points in Ω_1, and $((1-\alpha')/(J-1))\|\Omega_0\|^{-1}$ at all points in $\Omega_0 - \Omega_1$. Hence, there exists a scalar, $\alpha \in [0, 1]$, such that:

$$p(\vec{u} \mid D_1) = \alpha P_{\Omega_1}(\vec{u}) + (1-\alpha) P_{\Omega_0}(\vec{u}) = \begin{cases} \alpha \|\Omega_1\|^{-1} + (1-\alpha)\|\Omega_0\|^{-1} & \text{if } \vec{u} \in \Omega_1 \\ (1-\alpha)\|\Omega_0\|^{-1} & \text{if } \vec{u} \in \Omega_0 - \Omega_1, \end{cases} \quad (2)$$

where $P_{\Omega_1}(\vec{u})$ is the uniform distribution with support Ω_1. Equation (2) demonstrates that the posterior is a mixture of two uniform distributions supported by polyhedra. The scalar α is implicitly defined by equating Equations (1) and (2):

$$\frac{\alpha\|\Omega_1\|^{-1} + (1-\alpha)\|\Omega_0\|^{-1}}{(1-\alpha)\|\Omega_0\|^{-1}} = \frac{\alpha'(J-1)}{1-\alpha'} \Leftrightarrow \frac{\alpha\|\Omega_0\|/\|\Omega_1\| + (1-\alpha)}{(1-\alpha)} = \frac{\alpha'(J-1)}{1-\alpha'}.$$

The computation of α requires knowledge of $\|\Omega_0\|/\|\Omega_1\|$, which is the ratio of the measure of Ω_0 to the measure of Ω_1. For the choice-based polyhedral algorithm, we seek choice balance such that Ω_0 is divided into J collectively exhaustive equal-measure subpolyhedra; thus, the ratio $\|\Omega_0\|/\|\Omega_1\|$ is close to its average J. Hence, $\alpha \cong (J\alpha' - 1)/(J-1)$.

Let us next consider prior distributions that are defined by any mixture of uniform distributions supported by polyhedra, $\sum_{m=1}^{M} \omega_m P_{\Psi_m}(\vec{u})$. Defining Ω_1 as above and following the same argument, the posterior after the new choice question is proportional to:

$$P(D_1 \mid \vec{u})\left[\sum_{m=1}^{M} \omega_m P_{\Psi_m}(\vec{u})\right]$$
$$= \sum_{m=1}^{M} \omega_m (P(D_1 \mid \vec{u}) P_{\Psi_m}(\vec{u}))$$
$$= \sum_{m=1}^{M} \omega_m (\alpha_m P_{\Psi_m \cap \Omega_1}(\vec{u}) + (1-\alpha_m) P_{\Psi_m \cap \Omega_0}(\vec{u})),$$

which is also a mixture of uniform distributions supported by polyhedra.

Finally, we generalize to q questions. Let S_q be the set of all subsets of $\{1, 2, \ldots, q\}$. For a subset s of S_q, let $\Omega_s = \{\vec{u}\colon X_s \vec{u} \geq \vec{0}, \vec{u} \geq \vec{0}, \vec{e}'\vec{u} = 100\}$ be the polyhedron consistent with the choice questions contained in s (recall that X_s encodes the constraints implied by the answers to the questions in s). Let $w_s = \alpha^{|s|}(1-\alpha)^{q-|s|}$ for all nonempty Ω_s, where $|s|$ denotes the number of elements in subset s and $\alpha = (J\alpha' - 1)/(J-1)$. The posterior after q questions is a mixture of uniform distributions supported by the polyhedra $\{\Omega_s \cap \Psi_m\}_{s \in S,\, m \in \{1,\ldots,M\}}$. We approximate the weights as follows:[4]

$$P_q(\vec{u}) = \sum_{m=1}^{M} \sum_{s \in S_q} \omega_m w_s P_{\Omega_s \cap \Psi_m}(\vec{u}). \tag{3}$$

We denote question selection and estimation based on this posterior distribution as "polyhedral with error-modeling and informative priors." We also consider the following special cases in the simulations and field experiment:

- "polyhedral without error-modeling and with uniform priors" (as in THS): $\alpha = 1$, prior: $P_{\Omega_0}(\vec{u})$, posterior: $P_q(\vec{u}) = P_{\Omega_{\{1,\ldots,q\}}}(\vec{u})$;
- "polyhedral with error-modeling and with uniform priors:" $\alpha < 1$, prior: $P_{\Omega_0}(\vec{u})$, posterior: $P_q(\vec{u}) = \sum_{s \in S_q} w_s P_{\Omega_s}(\vec{u})$;
- "polyhedral without error modeling and with informative prior:" $\alpha = 1$, prior: $\sum_{m=1}^{M} \omega_m P_{\Psi_m}(\vec{u})$, posterior: $P_q(\vec{u}) = \sum_{m=1}^{M} \omega_m P_{\Omega_{\{1,\ldots,q\}} \cap \Psi_m}(\vec{u})$.

[4] We set $\omega_m w_s$ to zero if $\Omega_s \cap \Psi_m = \varnothing$ and normalize the weights to sum to one.

Selecting Questions and Estimating Partworths with Mixtures of Distributions

In the deterministic algorithm, THS select questions based on the analytic center and longest axes of a single polyhedron. This is a well-defined problem. For the probabilistic algorithm we must work with $P_q(\vec{u})$, which is a mixture of uniform distributions supported by polyhedra. To implement choice balance and postchoice symmetry, we must compute the analytic center and longest axes of polyhedral mixtures. Fortunately, the analytic center may simply be replaced with the appropriate mixture of the analytic centers of the polyhedra in the mixture. However, computing the longest axes poses a conceptual and technical challenge.

Longest Axes of a Mixture of Polyhedra

The longest axis of a mixture of polyhedra should summarize the directions of the longest axes of the polyhedra in the mixture and do so according to the weights of the mixture. Let $\vec{v}_{sm}$ be the longest axis of the polyhedron $\Omega_s \cap \Psi_m$ (see Equation (3)), and $\omega_m w_s$ be the corresponding weight. We seek the vector $\vec{v}^*$ that maximizes the weighted norm of the projections of $\vec{v}_{sm}$ on $\vec{v}^*$. Thus, the longest axis is the solution to the following mathematical program:

$$\text{(OPT1)} \quad \vec{v}_q^* = \arg\max_{\vec{v}} \sum_{m=1}^{M} \sum_{s \in S_q} \omega_m w_s (\vec{v}_{sm}^T \vec{v})^2.$$

Fortunately, OPT1 has a known solution. Define V as the matrix obtained by stacking the transposed longest axes, $\vec{v}_{sm}^T$. Define Π as the diagonal matrix with elements equal to the weights $\{\omega_m w_s\}$. We rewrite OPT1 in matrix form as follows:

$$\sum_{m=1}^{M} \sum_{s \in S_q} \omega_m w_s (\vec{v}_{sm}^T \vec{v})^2 = \vec{v}' V' \Pi V \vec{v}.$$

OPT1 is now a standard optimization problem that is analogous to factor analysis: $\vec{v}_q^*$ is the eigenvector associated with the largest eigenvalue of $V'\Pi V$. The matrix is symmetric and positive definite; hence, its eigenvalues are all real and nonnegative. The second-longest axis is associated with the second eigenvalue, etc. Because the axes are eigenvectors, they are guaranteed to be orthogonal.

Practical Implementation

While mixtures of polyhedra can approximate almost any distribution, there are practical considerations. Not only do the population priors grow exponentially with the number of features (2^F), but the number of subsets S_q in Equation (3) grows exponentially with the number of questions (2^q). For small q and small F, computation can be done quickly. Choice-based questions can be selected in less than a second, such that

respondents do not notice any delay. However, for large q or large F the delay can exceed a second (e.g., $2^{16} = 65{,}536$).

We take three steps to reduce computation time. First, the set of polyhedra in the posterior mixture after q questions, $\{P_{\Omega_s \cap \Psi_m}\}_{m=1,\ldots,M;\, s \in S_q}$, is a subset of the polyhedra in the posterior mixture after $q+1$ questions, $\{P_{\Omega_s \cap \Psi_m}\}_{m=1,\ldots,M;\, s \in S_{q+1}}$ (this follows from the fact that $S_q \subset S_{q+1}$). By saving, rather than recomputing, the longest axes and analytic center, we reduce computation time substantially. Second, one of the time-consuming steps in polyhedral methods is finding a feasible point in $\Omega_s \cap \Psi_m$. If a point is feasible in $\Omega_{\{1,2,\ldots,q\}} \cap \Psi_m$, then it is feasible for all $\Omega_s \cap \Psi_m$, $s \in S_q$. By reusing feasible points, we also reduce computational time substantially. Third, as the number of questions grows large, we sort the weights $\omega_m w_s$ in decreasing order and apply the algorithm to subsets corresponding to the largest weights, doing so until a preset time limit is reached. In our empirical work, that time limit is one second. In simulation, we use 10 seconds. Exploratory work suggests that these time limits provide excellent performance. However, all empirical and simulation results reported in this paper can be considered conservative and might improve slightly with faster computers and more efficient codes/programming/compilers.

5. Monte Carlo Simulations

Modeling response error and informative priors promises to enhance the accuracy of choice-based polyhedral question selection. However, both extensions increase complexity and could result in overfitting the data. To evaluate performance, we turn to complementary testing tools, both synthetic and empirical data. We use Monte Carlo simulations to study the potential of the methods by investigating the range of performance in a variety of relevant domains. With synthetic data we know the "truth" and can compare estimates to that benchmark. We use the field experiments to test practical implementation in a realistic setting. We do not know the true values of the partworths, and so must use predictive ability as a surrogate.

Experimental Design for the Monte Carlo Simulations

We use a $2 \times 2 \times 6 \times 4$ design for the Monte Carlo simulations. We simulate the respondents with a 2×2 subdesign that is becoming standard—allowing for two levels of response accuracy and two levels of respondent heterogeneity (Arora and Huber 2001, Evgeniou et al. 2005, Toubia et al. 2004). The Arora-Huber design, as modified by THS, uses four features at four levels each to ensure complete aggregate customization and orthogonal designs. Partworths are drawn from normal distributions with means, $\bar{\bar{u}}$, and variances, σ_β^2. The four levels of each partworth have means $[-\bar{\beta}, -\bar{\beta}/3, \bar{\beta}/3, \bar{\beta}]$. Higher values of $\bar{\beta}$ imply higher response accuracy. Higher values of σ_β^2 imply greater heterogeneity. We used the standard values of $\bar{\beta} = 1$ for the "low-accuracy" case and $\bar{\beta} = 3$ for the "high-accuracy" case with $\sigma_\beta^2 = \bar{\beta}$ for "low heterogeneity" and $\sigma_\beta^2 = 3\bar{\beta}$ for "high heterogeneity."

For each respondent we simulate six question-selection methods:

- random
- orthogonal
- aggregate customization (Huber and Zwerina 1996, Arora and Huber 2001)
- deterministic polyhedral (as in THS)
- probabilistic polyhedral with error modeling and uniform priors
- probabilistic polyhedral with error modeling and informative prior (prior approximates a normal distribution—see §3 and Appendix A3 for details).

The six question-selection methods are crossed with four estimation methods:

- hierarchical Bayes (HB) with normal priors
- deterministic analytic-center estimation (AC)
- analytic-center estimation with error modeling and uniform prior (ACe)
- analytic-center estimation with error modeling and informative prior (ACe+i) (prior approximates a normal distribution).

Simulated Environment

Aggregate customization uses relabeling and swapping to improve utility balance in choice-based questions and requires an estimate of the population partworth means. Polyhedral question selection with error modeling requires an estimate of α'. This estimate is derived from the same population estimates (see Appendix A1 for details). Following Arora and Huber (2001), we assume perfect pretest information. This should not affect the relative comparison of aggregate customization and probabilistic polyhedral question selection. Naturally, no such assumption is made in the empirical tests. Likewise, to investigate the impact of informative priors (relative to no priors), we use a rough approximation (four polyhedra) to the true prior distribution.[5] All simulation results are interpreted in light of these assumptions.

We seek to afford the estimation benchmark methods the strongest possible performance. Evgeniou

[5] We use normal priors rather than population priors on feature importances, because the latter would require that we deviate from the standard simulation design, thus reducing our ability to compare our results to previously published papers. Population priors only lead to differences when the average importances vary among the four simulated features. This does not happen in the standard simulation design, but is likely in our empirical test.

Table 1 Monte Carlo Simulation Results

Magnitude (accuracy)	Heterogeneity	Question selection method	RMSE HB[1]	RMSE AC	RMSE ACe	RMSE ACe+i
Low	High	Random	0.669	0.932	0.808	0.728
		Orthogonal	0.644	0.902	0.705	0.658
		Aggregate customization	0.610	0.916	0.733	0.670
		Deterministic polyhedral	0.599	0.778	**0.648***	0.614
		Probabilistic polyhedral w/ error modeling and uniform priors	**0.586***	**0.759***	**0.645***	**0.601***
		Probabilistic polyhedral w/ error modeling and informative priors	**0.584***	**0.767***	**0.645***	**0.598***
Low	Low	Random	0.645	0.963	0.824	0.653
		Orthogonal	**0.604***	0.913	**0.714***	0.606
		Aggregate customization	**0.597***	0.983	0.786	0.627
		Deterministic polyhedral	0.630	0.877	**0.708***	0.607
		Probabilistic polyhedral w/ error modeling and uniform priors	0.612	**0.837***	**0.713***	**0.603***
		Probabilistic polyhedral w/ error modeling and informative priors	**0.595***	**0.845***	**0.707***	**0.596***
High	High	Random	0.583	0.887	0.850	0.662
		Orthogonal	0.562	0.939	0.729	0.586
		Aggregate customization	0.514	1.026	0.785	0.613
		Deterministic polyhedral	0.497	0.680	**0.613***	0.528
		Probabilistic polyhedral w/ error modeling and uniform priors	0.487	0.665	**0.620***	0.521
		Probabilistic polyhedral w/ error modeling and informative priors	**0.448***	**0.633***	**0.611***	**0.476***
High	Low	Random	0.489	0.903	0.861	0.580
		Orthogonal	0.450	0.959	0.746	0.474
		Aggregate customization	0.404	1.057	0.815	0.499
		Deterministic polyhedral	0.441	0.702	**0.623***	0.468
		Probabilistic polyhedral w/ error modeling and uniform priors	0.438	**0.677***	**0.633***	0.480
		Probabilistic polyhedral w/ error modeling and informative priors	**0.392***	**0.671***	0.648	**0.422***

Notes. RMSE, lower is better.

*Best, or not significantly different from best, at $p \leq 0.05$ within magnitude × heterogeneity × estimation condition.

[1]HB = hierarchical Bayes estimation, AC = deterministic analytic-center estimation, ACe = analytic-center estimation with error modeling and uniform priors, ACe+i = analytic-center estimation with error modeling and informative priors.

et al. (2005) demonstrate that HB performs better if for each respondent we use rejection sampling (Allenby et al. 1995) to constrain the HB estimates so that the partworth of the lowest level of each feature is also the smallest partworth for that feature.[6] We adopt this procedure for HB in both the simulation and the field experiments.

Because polyhedral methods are designed for short Web-based questionnaires, we test designs of eight questions, choosing randomly for orthogonal and aggregate customization as in THS. For comparison to previously published simulations we report root mean squared error (RMSE) after normalizing the true and the estimated partworths so that their absolute values sum to the number of parameters and so that their values sum to zero for each feature. This enables us to interpret the RMSEs as a percentage of the mean (absolute) partworths. The simulations in Table 1 are based on now-standard 10 sets of 100 respondents. This is not a computational constraint; the field tests are based on larger samples.

[6] Constraints in estimation are used in other areas of marketing as well (see, for example, Ailawadi et al. 2005).

Results and Interpretation of Synthetic Data Experiments

Question Selection. Taking response errors into account and using informative priors appears to have the potential to improve question selection. At least one of the two modifications in polyhedral question selection is best or tied for best in all 16 accuracy × heterogeneity × estimation experimental cells. Probabilistic polyhedral question selection with error modeling and uniform priors is at least as good as the deterministic algorithm in every cell and significantly better in 9 of the 16 cells. Probabilistic polyhedral

question selection with error modeling and informative priors is at least as good as the deterministic algorithm in 15 of the 16 cells and significantly better in 12 of the 16 cells. The field experiment will test whether such improvements are sustained in practical implementations.

Estimation. Taking response errors into account and using informative priors also appear to have the potential to improve polyhedral estimation. At least one of the two improvements is better than deterministic analytic-center estimation in all accuracy × heterogeneity experimental cells. Informative priors appear to provide the greater improvement. However, the hierarchical Bayes estimates (HB) are still significantly better in three of the four accuracy × heterogeneity cells. The only exception is the low-accuracy, low-heterogeneity cell in which ACe+i is statistically tied with HB. These results are consistent with the simulations of Evgeniou et al. (2005).

In summary, our simulations suggest that incorporating response error and/or informative priors into polyhedral question selection is likely to enhance accuracy in empirical applications. Analytic-center estimation is also improved, but hierarchical Bayes is likely to remain the best estimation method in most application domains for choice-based questions.

6. Empirical Application and Test

Managerial Context

Traditional cork closures have dominated the wine industry for hundreds of years, but each year 5%–15% of all bottled wine is tainted due to poor-quality closures. Cork closures result in brand-name erosion and millions of dollars in lost revenue when consumers attribute the poor quality to the winery rather than the closure. As an alternative to cork closures, the wine industry developed Stelvins, a screw-cap/twist-off closure for mid-to-high-priced wines. Stelvins eliminate cork taint and other malodorous flavors, eliminate wine oxidation that leads to rapid aging, and minimize loss of fruit flavors due to air leakage. Stelvins provide "consistent, reliable, aging characteristics, showing the wine's development as the winemaker intended (Courtney 2001)."

Although Stelvins have been available for almost 50 years, and in Australia and New Zealand sales of premium wines with Stelvins now outnumber the sales of premium wines with corks, Stelvins are rarely used in the United States. To explore strategies for a U.S. introduction of Stelvins, a Napa Valley-based closure manufacturer and cooperating U.S. wineries asked us to determine preferences of leading-edge wine customers in the United States, New Zealand, and Australia.

Experimental Design

In exchange for gathering these data, the sponsors agreed to set up the application as an experimental design. Each respondent completed two sequential, rotated, choice-based conjoint analysis tasks separated by a series of "memory-cleansing" questions. The advantage of this experimental design is the increased power due to methodological comparisons within respondents.

We recruited 2,255 leading-edge wine consumers from the United States, Australia, and New Zealand (late 2004). Respondents were subscribers to wine-related e-newsletters (*WineX* and *WineBrats* in the United States, *Vine Cellars* in Australia and New Zealand) and could be expected to be knowledgeable about fine wines. They were likely to be leading-edge consumers. As a check, 80% of the respondents scored 15 or higher on a 21-point involvement scale (Lockshin et al. 2001). We obtained 245 respondents from a first U.S. panel, 958 from a second U.S. panel, 667 from Australia, and 385 from New Zealand. As is typical in managerial applications, we did not have total control over the assignment of respondents to treatments, although there was no reason to believe that there were any systematic biases within any of the countries.

Managerially, the sponsors were interested in the trade-offs that the consumers would make between wine closures and other features of wine. The conjoint design included five features at four levels each:

- closure type: traditional cork, synthetic cork, Metacork™,[7] Stelvin (screw cap);
- type of wine: dry white, aromatic white, dry red, blush red;
- origin: Australia/New Zealand, France, Sonoma/Napa, Chile/Argentina;
- vintner: small boutique, midsize regionally known winery, large nationally recognized winery, international conglomerate winery;
- price range:[8] four levels in the respondents' currency (e.g., Australian dollars).

The features in the conjoint design were introduced to respondents through self-explicated importance questions.[9] Figure 4a shows the closure types

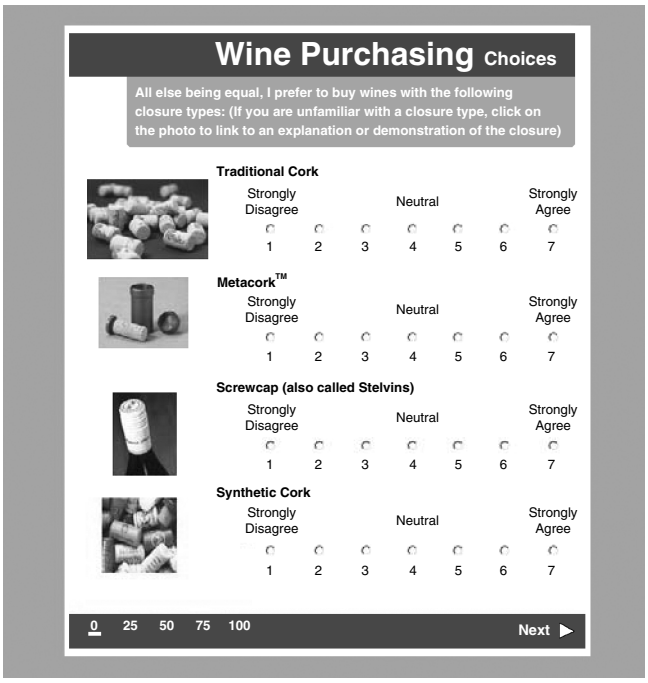
[7] A MetaCork™ "combines an integrated corkscrew, a drip-resistant pour feature, and a reseal cap." (www.metacork.com, viewed 2006)

[8] Based on pretests, respondents felt they could best evaluate the choices among wines if price was specified as a range. This is sufficient for relative methodological comparisons and the study of the impact of consumer preferences for Stelvin closures.

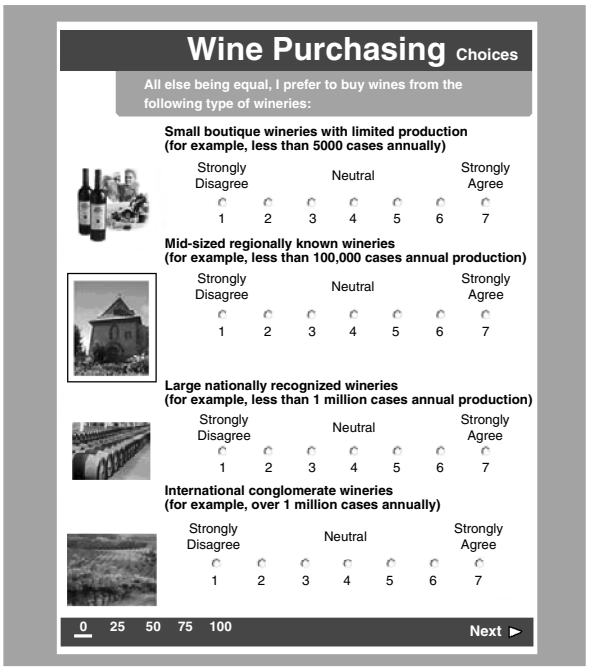
[9] These answers allowed us to identify the lowest level of each feature. This information was used in adaptive question selection and by all the estimation methods, including HB (see previous section). The self-explicated information was used in adaptive question selection and in estimation in order to avoid endogeneity and/or violations of the likelihood principle (Liu et al. 2006).

Figure 4 Example Screenshots from the Wine-Closure Preference Study

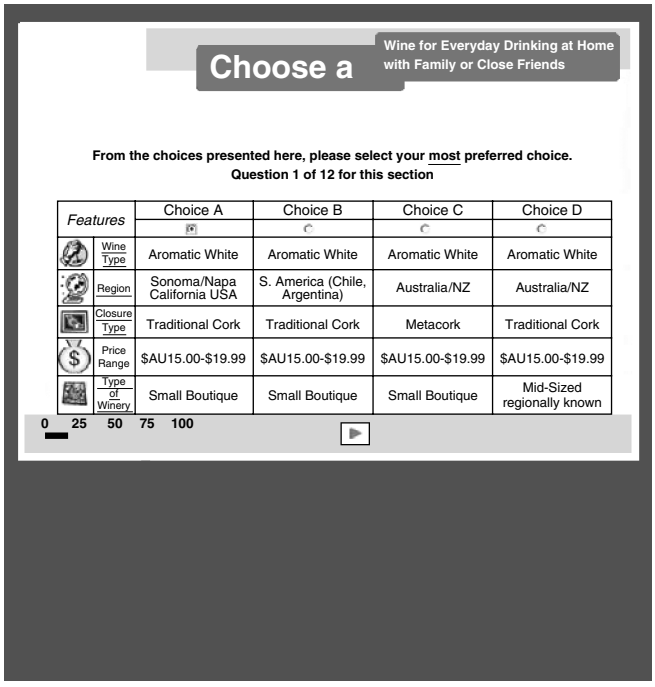
(a) Wine closures



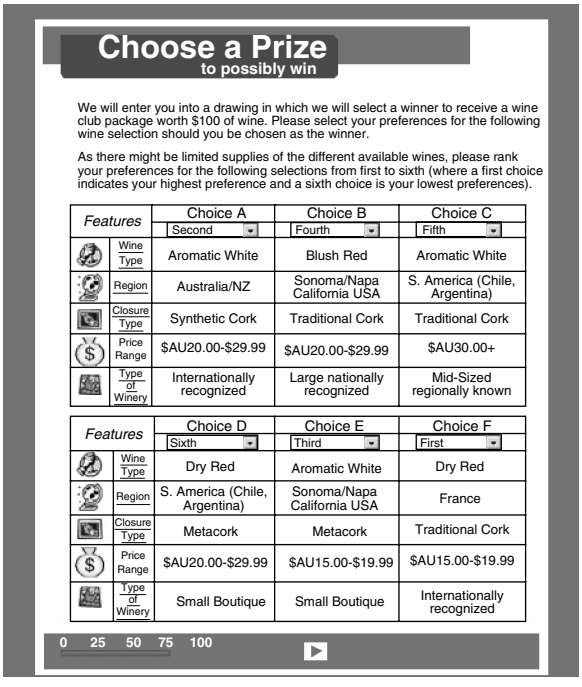
(b) Another feature (winery type)



(c) Choice-based questions



(d) Validation choice questions



and Figure 4b another feature (vintners). Respondents were then asked two sets of 12 choice-based questions as illustrated by Figure 4c. The first 10 questions of each set were designed by a different method (the order was rotated). The last two questions were randomly selected holdouts. Finally, after additional filler tasks, respondents were entered into a lottery with a 1 in 200 chance of winning a case of wine worth $100. Respondents were asked to rank six cases of wine and were told that they would receive their first choice if it was available. Otherwise, they would receive their second choice, etc. All bottles within a case were the

same (for example, if the wine costs \$20, they received five bottles).[10] The six wine profiles were randomly chosen from a 16-profile orthogonal design. This last task, designed with a different look and feel from the conjoint tasks (Figure 4d), serves as a validation.

We note that this study is the first empirical test of the predictive ability of choice-based polyhedral methods. Toubia et al. (2003) report on metric paired comparisons for laptop computer bags, and THS report on convergence and choice balance for an executive education study.

Comparisons of Question-Selection Methods: Experimental Design

We tested the following four question-selection methods:[11]

- orthogonal design;
- aggregate customization;
- deterministic polyhedral (THS);
- polyhedral with error modeling and uniform priors ("probabilistic polyhedral").

We chose the four methods carefully both to test the new probabilistic polyhedral method and to explore two fundamental characteristics. (1) Adaptation: Both deterministic and probabilistic polyhedral methods adapt questions within the respondent; aggregate customization and orthogonal designs do not. (2) Pretest information: aggregate customization and probabilistic polyhedral methods require pretest information to set "tuning" parameters; orthogonal designs and deterministic polyhedral methods do not. The pretest information was obtained from an HB analysis of 66 respondents who answered questions based on an orthogonal design. In order not to confound these two characteristics, we tested these methods in two pairs: orthogonal versus deterministic polyhedral, and aggregate customization versus probabilistic polyhedral.[12] We did not test random question selection because prior research suggests that aggregate customization and orthogonal design are stronger benchmarks. We leave tests of informative priors in question selection to future research. We test informative priors for estimation (see below).

Comparisons of Estimation Methods

Each of the estimation methods is compatible with all of the question-selection methods enabling us to make comparisons within respondent. The methods we tested were:[13]

- hierarchical Bayes estimation (HB);
- deterministic analytic-center estimation (AC);
- analytic-center estimation without error modeling and with informative population priors (ACi);
- analytic-center estimation with error modeling and with uniform priors (ACe).

Because partworth values might vary by panel and treatment (and some methods shrink estimates to the mean or median), we apply all methods within panel and treatment.

We begin with estimation and then move to our primary focus: question selection. Table 2 summarizes the comparisons of estimation methods for the validation task by reporting the correlation (averaged across respondents and question selection methods) between the predicted and observed rankings of the six wines in the validation task.[14]

We compared estimation methods statistically with a repeated-measures ANOVA, with performance as the dependent variable, two between-subject factors, panel (four levels) and question-selection comparison treatment (two levels); and two within-subject factors, estimation method (four levels) and a factor capturing whether the question selection is adaptive (two levels). We used contrast analysis to compare estimation methods. As predicted by the simulations, HB performs significantly better than the other estimation methods ($p < 0.01$).[15]

[10] Due to legal issues regarding alcohol as a prize, Australian respondents were not eligible to win real cases of wine. For these respondents the choice was hypothetical. Providing a reward with a fixed monetary value mitigates any wealth effect that might be present if we had endowed each respondent with money and given them the option of choosing among differently priced wines. The task remains incentive compatible as long as consumer utility is approximately linear in the number of bottles of wine over the range of the options available.

[11] The orthogonal and aggregate customization designs were the most D-efficient sets of 10 questions from a 16-question orthogonal design and from the corresponding aggregate customization design, respectively.

[12] Due to a programming error, 227 and 204 respondents from the Australian panel were assigned, respectively, to orthogonal versus aggregate customization and to deterministic polyhedral versus probabilistic polyhedral. Orthogonal versus aggregate customization revealed no difference. Probabilistic polyhedral performed better than deterministic polyhedral, although the sample size was insufficient to reach significance. Details are available from the authors.

[13] As is appropriate for an empirical test, we use our second form of informative priors in which the prior distributions are chosen to match the population medians. This method uses a mixture of 2^F polyhedra in the prior, where $F = 5$ is the number of features. Even with the computational efficiencies discussed earlier, it is not yet practical to apply both error modeling and population priors on our large data set. The former is exponential in the number of questions and the latter in the number of features. We leave development of faster heuristics to future research, noting that empirical results are thus conservative. If error modeling and population priors are separately promising, we might infer that their combination is also promising.

[14] Measuring performance by the proportion of respondents for whom the first choice in the validation task was correctly predicted or using holdout hit rate yields similar qualitative implications.

[15] The panel was significant at the 0.01 level and treatment was nonsignificant ($p = 0.42$). Adaptation is discussed below.

Table 2 Comparing Estimation Methods—Correlation with Choice

Experimental cell	Australian panel ($n = 667$)	New Zealand panel ($n = 385$)	First U.S. panel ($n = 245$)	Second U.S. panel ($n = 958$)	Average ($n = 2{,}255$)
Hierarchical Bayes (HB)	0.512	0.421	0.408	0.337	**0.411**
Analytic center w/o error modeling and w/ uniform prior (AC)	0.453	0.426	0.365	0.243	**0.350**
Analytic center w/ error modeling and w/ uniform priors (ACe)	0.465	0.420	0.353	0.279	**0.366**
Analytic center w/o error modeling and w/ informative priors (ACi)	0.475	0.434	0.389	0.245	**0.361**

Although the focus of this paper is on probabilistic polyhedral question selection, probabilistic analytic-center estimation is a byproduct of probabilistic question selection, and probabilistic analytic-center estimation improves predictions relative to deterministic analytic-center estimation. Including population priors (ACi) significantly improves performance ($p < 0.01$) compared to deterministic analytic-center estimation (AC). Including error modeling (ACe) improves performance as well, albeit not significantly ($p = 0.30$). (ACe is never significantly worse than AC, and is significantly better on the second U.S. panel.) The two improvements do not perform significantly differently ($p = 0.18$).

Empirical Comparison of Question-Selection Methods

Table 3 reports the average performance of the different question-selection methods (averaged across respondents and estimation methods). Due to circumstances beyond our control, all U.S. respondents were assigned to the "probabilistic polyhedral versus aggregate customization" condition. Notice that the average predictive ability varies between panels, with predictive ability significantly lower in the U.S. panels than in the Australian or New Zealand panels. While tempting, we cannot attribute these differences to across-country variation. Our panels were chosen from opt-in organizations of leading-edge wine users. These organizations might vary on other characteristics besides country of origin. Nonetheless, a future investigation of across-country differences in response quality would be interesting.

We examine significance with a repeated measures ANOVA on each experimental cell, with one between-subject factor, panel (four levels); and two within-subject factors, estimation method (four levels) and question-selection method (two levels).[16] Deterministic polyhedral question selection predicts significantly better than orthogonal question selection: across panels ($p < 0.07$) and within the Australian panel ($p < 0.05$ – ANOVA on the Australian panel only). Probabilistic polyhedral question selection performs significantly better than aggregate customization question selection: overall ($p < 0.06$), across the non-U.S. panels ($p < 0.02$), and in the New Zealand panel ($p < 0.05$). Probabilistic polyhedral question selection is better in three of the four panels and never significantly worse. While the results do not always obtain a significance level of 0.05, we can say, at minimum, that probabilistic polyhedral methods show promise.

In summary, the proposed probabilistic polyhedral question-selection methods improve correlations between predicted and actual choice in at least some situations. HB remains the best estimation method overall. Both ACe and ACi improve predictive ability relative to deterministic polyhedral methods (AC).

Substantive Results: Consumer Reactions to Stelvin Screw Caps for Fine Wine

Figure 5 reports the estimates of the average partworths for wine closures for leading-edge wine consumers in the United States, Australia, and New Zealand.[17] In Australia and New Zealand there is a slight preference for Stelvins over traditional cork closures. However, for the United States, corks are strongly preferred to Stelvins and closure type is a more important attribute. U.S. consumers even prefer MetaCorks™ and synthetic corks to Stelvins, whereas Australians and New Zealanders prefer Stelvins to these other nontraditional closures.

We also examine the importance of wine closures relative to other features. Figure 6 reports the average partworths for the type of wine and the origin of the wine. Preferences for the type of wine and the country of origin are roughly the same for U.S., New Zealand, and Australian consumers, with the exception of a home-country bias. (Detailed partworth values are available from the authors.)

The data suggest that for U.S. consumers, the relative importance of Stelvins versus corks (6.91) is

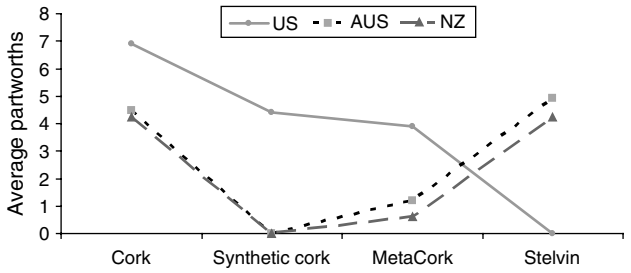
[16] Similar significance levels were obtained with an ANOVA similar to the ANOVA for the Table 2 data, with an additional between-subject factor capturing the experimental cell.

[17] The average partworths have been normalized such that the lowest level of each attribute has a partworth of 0 and the sum of the partworths across attributes is 100.

Table 3 Comparing Question Selection Methods—Correlation with Choice

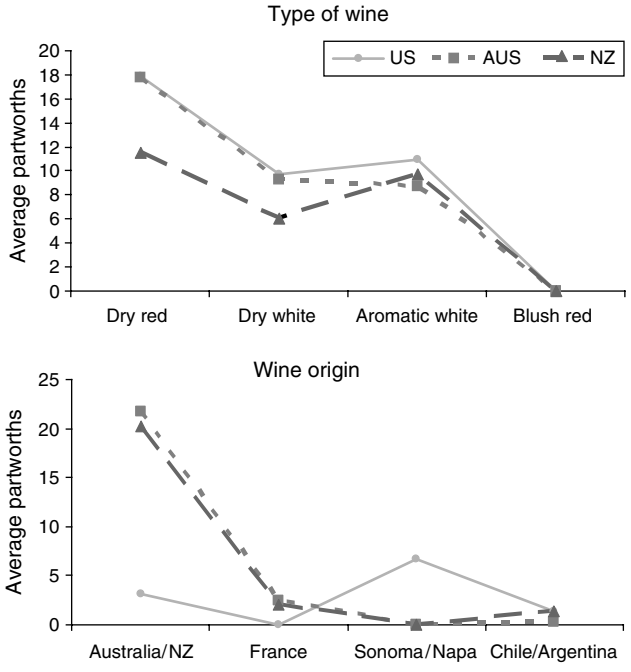
Experimental cell	Question-selection method	Australian panel	New Zealand panel	Average of Australia and NZ	First U.S. panel	Second U.S. panel	Average of U.S. panels
Deterministic polyhedral vs. orthogonal ($n = 527$)	Deterministic polyhedral	0.492	0.429	**0.468**	n/a	n/a	n/a
	Orthogonal	0.445	0.405	**0.430**	n/a	n/a	n/a
Probabilistic polyhedral vs. aggregate customization ($n = 1{,}728$)	Probabilistic polyhedral	0.498	0.457	**0.483**	0.373	0.282	**0.300**
	Aggregate customization	0.470	0.411	**0.449**	0.385	0.270	**0.293**

Figure 5 Average Partworths for Wine Closures in the U.S., Australia, and New Zealand Studies



comparable or less than the relative importance of wine type (dry red versus blush red, 17.88), region (United States versus France, 6.68), and type of winery (regional versus international, 8.36). At least initially, bottles with Stelvin closures will have to be offered at a discount in order to capture a significant market share. For example, for higher-priced wines, market simulations based on our estimates (averaged across question-selection methods) suggest that combining Stelvins with a $10 discount would allow capturing a 42.1% market share.[18] Based on these results, it appears that (1) there is current resistance to Stelvins among leading-edge U.S. consumers, (2) the importance of closure is not high relative to other features of wine, and (3) for current U.S. leading-edge consumers, a modest price discount might encourage the adoption of Stelvins.

Figure 6 Average Partworths for the Type of Wine and Wine Origin



Do the Managerial Recommendations Change Based on Question-Selection Method?

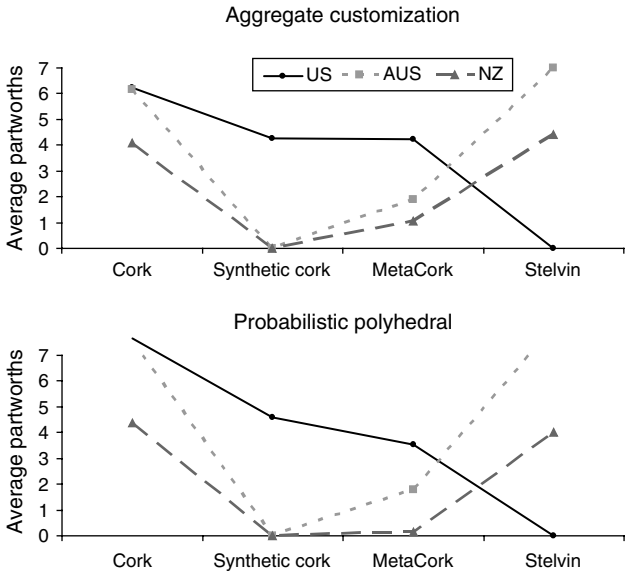
As an illustration, Figure 7 plots the estimates of the average partworths for wine closures based on the two question-selection methods in the U.S. panels. Comparing the two plots, we see subtle differences between methods. However, these plots only capture the average partworths across respondents, not the full distribution of partworths estimates. Moreover, without reference to the managerial context, it is difficult to intuit whether the estimates based on different question-selection methods imply differences in strategy. Thus, we examine the quantitative implications.

We begin by comparing the predicted response to a price discount of $10 on Stelvin closures. Partworth estimates based on aggregate customization questions suggest that a price discount of $10 on Stelvin closures would capture 44.3% of this premium wine market; partworth estimates based on the probabilistic polyhedral questions suggest lower market share of 39.8% ($p < 0.03$). Depending on the costs of marketing Stelvins, this lower reward might be the difference between a GO and a NO GO decision.

To gain further insight into whether or not partworth differences imply different managerial decisions, we draw on recent research by Belloni et al. (2005). Belloni et al. solve an optimal product design problem based on partworth data similar in structure to that collected here. Their design problem consists of selecting product features for the profiles in a product line in order to maximize profit (faced with a fixed set of competitors). Using Lagrangian relaxation with branch and bound, they identify the optimal product

[18] Results are approximate because the sponsors defined prices with ranges. We used the midpoint of each range in our share and profit calculations.

Figure 7 Average Partworths for the Wine Closures by Two Different Methods



line for relatively large numbers of features and customers. More importantly, they compare a variety of heuristics and demonstrate that simulated annealing (1) is feasible for reasonably sized problems and (2) achieves 100% of the optimum in their test problems (p. 20). We adopt their structure and modify their simulated annealing code to optimize a product line of wine profiles based on cost estimates obtained from wine experts. We assumed that the competitive products available to consumers were the set of all profitable profiles containing traditional corks. We assumed that each consumer purchased exactly one bottle.

Using Belloni et al.'s (2005) formulation, we developed optimal Stelvin-based product lines (consisting of 10 products) using (1) probabilistic polyhedral questions and (2) aggregate customization questions. Within this framework, the product line based on the partworths obtained by probabilistic polyhedral questions had three profiles in common with that designed based on the aggregate customization partworths. Another four pairs of profiles varied on one feature. Furthermore, if the partworths obtained from probabilistic polyhedral questions describe the market, the profit obtained with the polyhedral-based product line was 19.4% higher than the product line based on aggregate-customization partworths.[19]

[19] This calculation assumes that probabilistic polyhedral is the best estimate and is provided for illustration. Profit is guaranteed to be no worse by the principle of optimality. However, we find the magnitude of the difference—almost 20%—to be interesting, especially compared to the 2.1% difference due to optimization method found by Belloni et al. (2005, p. 28). For alternative methods of profitability comparisons, see Rust and Verhoef (2005).

7. Conclusions and Future Research

This paper focuses on improved methods for adaptive question selection in conjoint analysis. We nest deterministic polyhedral methods using *conjugate* classes of likelihood functions and prior distributions. Our probabilistic Bayesian framework overcomes prior weaknesses by enabling researchers to (1) take response error into account and (2) introduce informative priors. Simulation and empirical tests suggest that these improvements are promising. The wine-closure application is the first predictive test of choice-based polyhedral methods. For this application, individual adaptation of questions shows promise.

We close by noting limitations and avenues for future research. First, computation issues forced us to use approximations and to use prior distributions described by only few polyhedra. More efficient algorithms could be developed and the structure of the problem may be exploited further to alleviate this limitation. Second, analytic-center estimation continues to improve, but does not yet perform as well as hierarchical Bayes. Using the Bayesian interpretation of polyhedral question selection, we might derive a formal Bayesian-loss-function minimization that improves analytic-center estimation (Rossi and Allenby 2003). Third, as a first approximation we used the same value of α for all choice questions and all respondents. We might improve estimation if α is specified as a function of the difficulty of the choice questions and/or is allowed to vary over the number of questions (Liechty et al. 2005). Using the formulae in this paper, we might also specify a prior on α, conditional on the partworths and the choice set, and formulate a posterior given the observations.[20] Fourth, our simulations used a now-standard structure, but there remain interesting tests with nondiagonal covariance matrices and specifications where the average partworths vary. Finally, other approaches to handling response error may be developed using stochastic optimization (Spall 2003) or statistical learning theory (Evgeniou et al. 2005). Polyhedral methods remain a nascent technique that we hope will improve with future testing and future developments.

Acknowledgments

This research was supported by the MIT Sloan School of Management, MIT's Center for Innovation in Product Development, and Northeastern University's Institute for Global Innovation Management. The authors wish to thank Kevin Stanik and Andrew Rutkiewitz for their computer support, as well as Tom Atkin and Larry Lockshin with

[20] We would like to thank the AE for suggestions regarding the parameter α.

their wine industry expertise. Michael Yee has provided valuable comments that the authors gratefully acknowledge. They also thank Matthew Selove for suggesting the product-line optimization test.

Appendices. Derivations and Algorithms

A1. Computation of α', the Tuning Parameter for Error Modeling

We compute α' as follows:

- Estimate a population mean for the partworths, $\bar{\vec{u}}_{pop}$. In Monte Carlo simulations we use the true mean; in the empirical applications we use hierarchical Bayes estimates from the pretest subjects.
- Generate R random questions ($R = 100$ in the simulations and experiments) with logistic probabilities based on $\bar{\vec{u}}_{pop}$. The probability, α'_r, that a respondent chooses the maximum utility profile is the maximum logistic probability for that respondent on that question. Averaging over respondents gives an initial estimate, $\hat{\alpha}'_o$.
- Use $\bar{\vec{u}}_{pop}$ to simulate N respondents ($N = 100$ in the simulations and experiments) using $\alpha_o = (J\hat{\alpha}'_o - 1)/(J-1)$ for polyhedral question selection. Recompute $\hat{\alpha}'$ as above (assuming logistic probabilities).

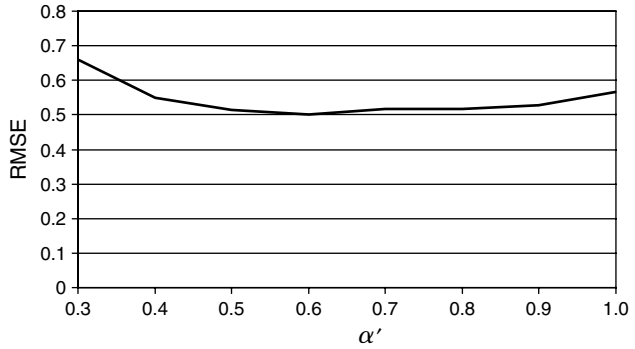
In theory, one might iterate these steps toward convergence; however, in practice we found that $\hat{\alpha}'$ was not sensitive to the initial estimate, $\hat{\alpha}_o$, used to generate the questions. Nonetheless, this means that our simulations are conservative.

A2. Sensitivity to α'

In the simulations, we purposefully base the choices of synthetic respondents on logistic probabilities. These simulated choices imply that α' varies by respondent and thus tests the sensitivity of our approximation that α' is constant across respondents. We study further the sensitivity with respect to α' using a simulation set up similar to that in §4. We simulate five sets of 100 synthetic respondents using magnitude and heterogeneity parameters equal to 2.0. Questions are selected with probabilistic polyhedral methods with error modeling and informative priors; the estimation method is ACe+i.

Figure A.1 suggests that predictive accuracy is flat for a fairly wide interval for α'. Fortunately, the values obtained using the above procedure always fell within this interval for both the simulations and the empirical tests. Predictions degrade only as α' is selected to imply almost random choice (toward 0.3) or almost no response error (toward 1.0).

Figure A.1 Sensitivity of Predictive Accuracy to α'



A3. Mixture Weights for Informative Priors (Mixtures of Normal Distributions)

Our goal is to approximate a normal distribution $N(\vec{u})$ by a mixture of uniform distributions supported by polyhedra:

$$\sum_{m=1}^{M} \omega_m P_{\Psi_m}(\vec{u}) + \left(1 - \sum_{m=1}^{M} \omega_m\right) P_{\Omega_0}(\vec{u}),$$

where $\Psi_m = \{\vec{u}\colon \hat{\vec{u}} - C_m \le \vec{u} \le \hat{\vec{u}} + C_m,\ \vec{u} \ge 0,\ \vec{e}'\vec{u} = 100\}$. In the Monte Carlo simulations, we used $M = 3$, with $C_1 = 5$, $C_2 = 10$, and $C_3 = 15$. Without loss of generality, assume that $C_1 < C_2 < \cdots < C_M$ so that the "boxes" used to approximate the normal distribution are of increasing sizes. The weights $\omega_1, \ldots, \omega_M$ are found by solving the following system of equations:

$$\begin{aligned}
&\text{Prob}(\vec{u} \in \Psi_1 \mid \vec{u} \sim N(\vec{u}))\\
&= \omega_1 + \omega_2 \text{Prob}(\vec{u} \in \Psi_1 \mid \vec{u} \sim P_{\Psi_2}) + \cdots + \omega_M \text{Prob}(\vec{u} \in \Psi_1 \mid \vec{u} \sim P_{\Psi_M})\\
&\quad + \left(1 - \sum_{m=1}^{M} \omega_m\right) P(\vec{u} \in \Psi_1 \mid \vec{u} \sim P_{\Omega_0}),\\
&\text{Prob}(\vec{u} \in \Psi_2 - \Psi_1 \mid \vec{u} \sim N(\vec{u}))\\
&= \omega_2 \text{Prob}(\vec{u} \in \Psi_2 - \Psi_1 \mid \vec{u} \sim P_{\Psi 2}) + \omega_3 \text{Prob}(\vec{u} \in \Psi_2 - \Psi_1 \mid \vec{u} \sim P_{\Psi_3})\\
&\quad + \cdots + \omega_M \text{Prob}(\vec{u} \in \Psi_2 - \Psi_1 \mid \vec{u} \sim P_{\Psi_M})\\
&\quad + \left(1 - \sum_{m=1}^{M} \omega_m\right) P(\vec{u} \in \Psi_2 - \Psi_1 \mid \vec{u} \sim P_{\Omega_0})\\
&\quad \cdots\\
&\text{Prob}(\vec{u} \in \Psi_M - \Psi_{M-1} \mid \vec{u} \sim N(\vec{u}))\\
&\quad = \omega_M \text{Prob}(\vec{u} \in \Psi_M - \Psi_{M-1} \mid \vec{u} \sim P_{\Psi_M})\\
&\qquad + \left(1 - \sum_{m=1}^{M} \omega_m\right) P(\vec{u} \in \Psi_M - \Psi_{M-1} \mid \vec{u} \sim P_{\Omega_0}).
\end{aligned}$$

In the Monte Carlo simulations the left-hand sides were approximated numerically by drawing 10,000 sets of parameters from $N(\vec{u})$, where $N(\vec{u})$ was the distribution used to generate the true partworths. $\text{Prob}(\vec{u} \in \Psi_m - \Psi_{m-1} \mid \vec{u} \sim P_{\Omega_0})$ was computed numerically by drawing 10,000 sets of parameters from P_{Ω_0}. We recognize that $\text{Prob}(\vec{u} \in \Psi_m - \Psi_{m-1} \mid \vec{u} \sim P_{\Psi_m})$ is equal to

$$\frac{\text{Prob}(\vec{u} \in \Psi_m - \Psi_{m-1} \mid \vec{u} \sim P_{\Omega_0})}{\text{Prob}(\vec{u} \in \Psi_m \mid \vec{u} \sim P_{\Omega_0})}.$$

The numerator and denominator were computed numerically.

A4. Incorporating Population Priors for Feature Importances

In our empirical application, all features have the same number of levels, so we define *importance* as the sum of the partworths for that feature (setting the lowest partworth to zero). Importance can also be defined as the difference between the highest and lowest partworth for a feature.

We have found that constructing priors based on constraints on the importances is more practical and intuitively appealing than using constraints on the partworths themselves.

Let m_f be the median of the importance of feature f based on individual-respondent estimates. Let $\hat{m}_f$ be the median importance based on $P_{\Omega_0}(\vec{u})$ and let $\hat{P}^u$ be the probability that the importance of feature f is smaller than m_f for $P_{\Omega_0}(\vec{u})$. Compute $\hat{P}^u$ numerically with 10,000 draws on Ω_0. If $m_f > \hat{m}_f$ ($\hat{P}^u > 0.5$), the constraint corresponding to feature f is that its importance is greater than m_f. This constraint is associated with the weight δ_f, such that $0.5 = \delta_f + (1-\delta_f).(1-\hat{P}^u)$. If $m_f \leq \hat{m}_f$ ($\hat{P}^u \leq 0.5$), the constraint corresponding to feature f is that its importance is less than m_f, and the corresponding weight is δ_f such that $0.5 = \delta_f + (1-\delta_f)\hat{P}^u$.

Let F be the number of features, and S_F be the set of subsets of all subsets of $\{1, 2, \ldots, F\}$. The prior distribution is $P(\vec{u}) = \sum_{s \in S_F} \omega_{m_s} P_{\Psi_s}(\vec{u})$, where

- $\omega_{m_s} = \prod_{f=1}^{F} \delta_f^{f \in s}(1-\delta_f)^{f \notin s}$ where $f \in s$ is 1 if f is in s and 0 otherwise; $f \notin s$ is its complement.
- Ψ_s is the polyhedron obtained from adding the constraints corresponding to the features $f \in s$ to the initial constraints defining Ω_0.

A5. Summary of Probabilistic Polyhedral Question Selection

1. Compute the weights for the probability mixture, $w_s = \alpha^{|s|}(1-\alpha)^{i-|s|}$, for all $s \in S_q$.

2. Compute the analytic center of the mixture, $\vec{\hat{u}}_q = \sum_{m=1}^{M} \sum_{s \in S_q} \omega_m w_s AC(\Omega_s \cap \Psi_m)$.

3. Approximate each polyhedron $\Omega_s \cap \Psi_m$ with an ellipsoid and compute the longest axis of the ellipsoid according to deterministic polyhedral methods (see THS for details).

4. Solve for the eigenvalues of $V'\Pi V$ and select the $J/2$ eigenvectors associated with the largest $J/2$ eigenvalues. (If J is odd, find the $(J+1)/2$ longest axes.) See the section "Longest Axes of a Mixture of Polyhedra" for details.

5. Find the intersections of the longest axes of the probability mixture with $\Omega_0 \Rightarrow \vec{u}_j$.

6. Find the J profiles by solving the knapsack problem, maximize $\vec{x}_j \vec{u}_j$ subject to $\vec{x}_j \hat{\vec{u}}_q \leq K$, where K is a randomly drawn constant. (See THS for details.)[21]

[21] In this paper, we drew K up to 30 times until all profiles were distinct. If all profiles are identical after 30 draws, it is likely that no further questions are needed and the questioning sequence stops. If after 30 draws there are only K' distinct solutions ($1 < K' < K$), these are presented to the respondent.

References

Ailawadi, K. L., P. K. Kopalle, S. A. Neslin. 2005. Predicting competitive response to a major policy change: Combining game-theoretic and empirical analyses. *Marketing Sci.* **24**(1) 12–24.

Allenby, G. M., N. Arora, J. L. Ginter. 1995. Incorporating prior knowledge into the analysis of conjoint studies. *J. Marketing Res.* **32**(May) 152–162.

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

Belloni, A., R. Freund, M. Selove, D. Simester. 2005. Optimizing product line designs: Efficient methods and comparisons. *Management Sci.* Forthcoming.

Courtney, S. 2001. Screwcap wine seals: Has Kiwi ingenuity gone too far? *Wine of the Week* (August). http://www.wineoftheweek.com/screwcaps/history.html.

Evgeniou, T., C. Boussious, G. Zacharia. 2005. Generalized robust conjoint estimation. *Marketing Sci.* **24**(3) 415–429.

Hauser, J. R., O. Toubia. 2005. The impact of utility balance and endogeneity in conjoint analysis. *Marketing Sci.* **24**(3) 498–507.

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

Kanninen, B. J. 2002. Optimal design for multinomial choice experiments. *J. Marketing Res.* **36**(May) 214–227.

Liechty, J. C., D. K. H. Fong, W. DeSarbo. 2005. Dynamic models incorporating individual heterogeneity: Utility evolution in conjoint analysis. *Marketing Sci.* **24**(2) 285–293.

Liu, Q., T. Otter, G. M. Allenby. 2006. Investigating endogeneity bias in marketing. *Marketing Sci.* **26**(5) 640–648.

Lockshin, L., P. Quester, T. Spawton. 2001. Segmentation by involvement or nationality for global retailing: A cross-national comparative study of wine shopping behavior. *J. Wine Res.* **12**(December) 223–236.

Rossi, P. E., G. M. Allenby. 2003. Bayesian statistics and marketing. *Marketing Sci.* **22**(3) 304–328.

Rust, R. T., P. C. Verhoef. 2005. Optimizing the marketing interventions mix in intermediate-term CRM. *Marketing Sci.* **24**(3) 477–489.

Spall, J. C. 2003. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Wiley-Interscience, Hoboken, NJ.

Toubia, O., J. R. Hauser, D. I. Simester. 2004. Polyhedral methods for adaptive choice-based conjoint analysis. *J. Marketing Res.* **41**(February) 116–131.

Toubia, O., D. Simester, J. R. Hauser, E. Dahan. 2003. Fast polyhedral adaptive conjoint estimation. *Marketing Sci.* **22**(3) 273–303.