Ellipsoidal Methods for Adaptive Choice-Based Conjoint Analysis

Denis Sauré
University of Chile, dsaure@dii.uchile.cl

Juan Pablo Vielma
MIT Sloan School of Management, jvielma@mit.edu

Questionnaires for adaptive choice-based conjoint analysis aim at minimizing some measure of the uncertainty associated with estimates of preference parameters (e.g. partworths). Bayesian approaches to conjoint analysis quantify this uncertainty with a multivariate distribution that is updated after the respondent answers. Unfortunately, this update often requires multidimensional integration, which effectively reduces the adaptive selection of questions to impractical enumeration. An alternative approach is the polyhedral method by Toubia et al. (2004), which quantifies the uncertainty through a (convex) polyhedron. The approach has a simple geometric interpretation, and allows for quick credibility-region updates and effective optimization-based heuristics for adaptive question-selection. However, its performance deteriorates with high *response-error* rates. Available adaptations to this method do not preserve all of the geometric simplicity and interpretability of the original approach. We show how, by using normal approximations to posterior distributions, one can include response-error in an approximate Bayesian approach that is as intuitive as the polyhedral approach, and allows the use of effective optimization-based techniques for adaptive question-selection. This *ellipsoidal* approach extends the effectiveness of the polyhedral approach to the high response-error setting and provides a simple geometric interpretation (from which the method derives its name) of Bayesian approaches. Our results precisely quantify the relationship between the most popular efficiency criterion and heuristic guidelines promoted in extant work. We illustrate the superiority of the ellipsoidal method through extensive numerical experiments.

Key words: conjoint analysis, geometric methods, Bayesian models, mixed-integer programming

1. Introduction

Conjoint analysis (Green and Srinivasan 1978) is a set of methods used in market research to elicit consumer preferences for products or services. Market researchers often envision products as vectors of attributes, or profiles; consumer preferences are thus characterized in terms of a "partworth" vector, which indicates how much a consumer values each particular attribute (in relative terms). In conjoint studies, representative consumers are asked a series of questions about their preferences; the data from their answers is used to estimate the parameters of an underlying model of consumer preferences, as well as estimates of the individual partworth vectors.

In choice-based conjoint analysis (Louviere et al. 2000), each question presents a consumer with a set of product profiles, among which the consumer is asked to choose the most preferred one.

The resemblance of this task to the actual environment consumers face in the marketplace has contributed to the popularity of choice-based methods.

Different questionnaires differ on the data collected from their application. Efficiency of a questionnaire is often measured in terms of the "precision" of the parameter estimates produced from the data it generates. A natural measure of the precision of these estimates is a function of the determinant of the covariance matrix of the estimates, which can be interpreted as function of the volume of a *credibility region* or *confidence set* around the estimate. Such a measure is usually denoted the *D-error* and minimizing it is often referred to as the *D-criterion* or seeking *D-efficiency*. McFadden (1974) shows that the D-error of a questionnaire depends on the underlying model parameters, thus indicating that efficient questionnaire design should include some form of sequential process that adjusts questions as information on said parameters is inferred from respondents' answers (Silvey 1980). Effectively, choice-based conjoint methods have evolved from static questionnaire designs that ask the same set of questions to all respondents, to heterogeneous designs where questions are designed using data from pre-tests and not every respondent answers the same set of questions.

In the last decade, advances in technology have allowed the development and implementation of *adaptive* designs that adapt questions within a questionnaire based on the answers of the respondent to previous questions. Toubia et al. (2004) in particular develops the polyhedral method for choice-based conjoint analysis. The method allows for fast updates of the credibility region, which is envisioned as a polyhedron, and selects questions adaptively using optimization-based methods. Unfortunately, the method does not perform well in settings with high response-error rates,[1] and while many extensions addressing the issue have been proposed in the literature, none of them maintain the simple and geometric intuition of the original method, nor its computational implementability.

Objectives and assumptions. In this paper we develop an adaptive questionnaire method for choice-based conjoint analysis that aims at *maximizing the precision of the individual partworth estimates*. For this, we envision adaptive question-selection as a sequential decision-making problem under uncertainty and model the latter in a Bayesian framework. In this context, we model the question-selection problem using dynamic programming and use *moment-matching* or *expectation-propagation* techniques (e.g. (Gelman et al. 2013, Section 13.8)) to approximate the posterior distribution given the response to a question. We then combine this approximation with mixed integer programming (MIP) techniques to formulate the optimal selection of the next question. We finally connect all these elements into a *near-optimal one-step look-ahead moment-matching approximate Bayesian policy* we denote the *ellipsoidal method* due to its natural geometric interpretation. In formulating the adaptive questionnaire design problem in a Bayesian framework we assume: *i*) a binary encoding of the product profiles, so that possible constraints on product design are explicitly considered when selecting questions; that *ii*) consumer's choice is driven by a *mixed logit* model, where we assume that a (multivariate normal) prior distribution for the partworth vector is available; and that *iii*) the number of questions in each questionnaire is fixed and given upfront, and that on each question a respondent is presented with *two* product profiles, with no implicit outside alternative.

The first assumption above is made to simplify the exposition (recall that D-error minimization is invariant to recoding of attributes, (Arora and Huber 2001)). The assumption on the specific choice model is made to facilitate the discussion and is quite common in the literature on questionnaire design; all methods in this paper extend directly to the case of alternative specifications for the idiosyncratic shocks to utility. The assumption of a prior distribution on the partworth vector is also common in the literature, with the most studied case being that of a multivariate normal distribution (see, e.g. Sándor and Wedel (2005)): see further comments and discussion in Section 4.4. Finally, while most theoretical results in the paper are developed for the case of two profiles per question, our methods extend to the case of more profiles per question and presence of an explicit outside alternative; see the discussion in Section 7.

Main Contributions. A first contribution lies in formulating the adaptive questionnaire design problem as a sequential decision-making problem under uncertainty. In particular, our formulation *explicitly* considers the objective of reducing the D-error of the partworth estimates. This provides a clear guideline for comparing adaptive methods.

The second and arguably more significant contribution lies in proposing a method that approximates closely a one-step look-ahead optimal policy, and whose question-selection does not rely on impractical enumeration. This method arises from applying an approximate dynamic programming approach to finding the "optimal" one-step look-ahead questionnaire policy. In designing the method we make two approximations. First, we follow a moment-matching approximate Bayesian approach and approximate the posterior partworth distribution with a multivariate normal distribution whose parameters are chosen so as to match those of the actual posterior distribution. We show that implementing such an approximation only requires one-dimensional integration, independent of the dimension of the partworth vector, and therefore it can be performed efficiently (see Proposition 1). The second approximation relates to efficient question-selection: we show that the expected posterior D-error (when the prior distribution is a multivariate normal) is a function of only two scalars that are quadratic functions of the encoding of the profiles in a question. In particular, we show that such an expected D-error depends on: *i*) the expected utility imbalance between the profiles in the question; and *ii*) the variance of the posterior along the direction of the question (see Proposition 2). This allows us to model optimal question-selection as a mixed integer programming (MIP) problem, which can be solved effectively using state-of-the-art MIP solvers.

Moreover, we show that such an approach to question-selection extends to alternative variance measures such as those based on asymptotic properties of the Fisher information matrix considered in extant work (see Section 5.4).

The third contribution comes from interpreting the proposed method's geometry and comparing it to that of existing methods. Following Toubia et al. (2004) one can envision adaptive methods as operating directly over credibility regions of the partworth vector (in our method, such credibility regions are ellipsoids, and are updated via posterior computation). This interpretation, from which the method derives its name, allows us to derive insight about the key principles behind optimal question-selection, and to interpret them in terms of key guidelines driving question-selection in previous work. In particular, in the polyhedral method questions are envisioned as hyperplanes cutting through the (polyhedral) credibility region, and question-selection follows two main guidelines; *i*) choice-balance; and *ii*) post-choice symmetry. Both guidelines aim at minimizing D-error; the first one is derived by the influential concept of *utility balance* (Huber and Zwerina 1996) in non-adaptive questionnaire design; the second aims at obtaining "symmetrical" credibility regions, which can be linked to reducing the volume of the expected credibility region. Prior work follow these guidelines, often prioritizing one over the other, but do not identify a precise trade-off between them (e.g. Toubia et al. (2004), Bertsimas and O'Hair (2013)). In contrast, we characterize the precise balance between choice-balance and post-choice symmetry needed to minimize the D-error. In particular, our analysis reveals that the D-error is a non-trivial, but precisely characterized function of choice-balance and post-choice symmetry. Furthermore, this function can be efficiently computed numerically through one-dimensional integration. This result reinforces the theoretical justification of the choice-balance and post-choice symmetry guidelines advocated for in Toubia et al. (2004) (see Proposition 2).

Finally, we present extensive numerical experiments showing that: (*i*) question-selection can be performed in real-time with no noticeable delay between questions, thus providing strong evidence of the implementability of the method; and that (*ii*) the ellipsoidal method outperforms other geometric-based adaptive questionnaire design methods in various criteria. With regard to (*i*), although our implementation relies on MIP and numerical integration, efficient coding allows us to solve MIP models and to approximate posterior distributions almost instantaneously. With regard to (*ii*), our results show that the ellipsoidal method consistently outperforms the benchmark in terms of minimizing D-error. Our experiments also show that the ellipsoidal method provides a better estimate of the partworth vector, both with respect to the distance between the true and estimated vectors and with respect to out-of-sample hit-rates. Furthermore, this advantage holds for the native estimators of each method and for individual and Hierarchical Bayes estimators computed offline numerically. Finally, this advantage is also preserved when the prior distribution used by the methods is not centered at the true population mean. The complete source code of the method implementations used for these experiments can be found at (Sauré and Vielma 2017).

Organization of the paper. The next section provides an overview on the relevant literature on adaptive conjoint analysis. In Section 3 we present a Bayesian framework for optimal questionnaire design in choice-based adaptive conjoint analysis. Then, in Section 4 we interpret existing geometric methods within the proposed framework, and discuss the use of prior information. In Section 5 we present the proposed method for near-optimal question-selection, and provide the geometric interpretation. Finally, Section 6 illustrates the performance of the ellipsoidal method via a comprehensive set of numerical experiments, and Section 7 presents our final remarks. Proofs and complementary computational results are included in Appendix A.

2. Literature Review

Static questionnaire design and efficiency criteria. Efficiency of a questionnaire design is commonly measured in terms of the "precision" of the parameter estimates, which are calibrated with the data collected through the questionnaire. This precision is typically associated with some function of the covariance matrix of the parameter estimates. For the case of choice experiments under the random utility framework, McFadden (1974) shows that maximum likelihood estimates are asymptotically normal, with a covariance matrix proportional to the inverse of the Fisher information matrix. For this reason, and while various efficiency criteria have been proposed (see, e.g. Kuhfeld et al. (1994) and Yu et al. (2012) for overviews and discussion), the literature has largely focused on minimizing the determinant of the inverse of the Fisher information matrix, an approach known as D-criterion or D-efficiency. The use of this criterion is often justified as the volume of an credibility region for a multivariate normal distribution is proportional to the square root of the determinant of its covariance matrix (see Segall (2010) for a discussion in a Bayesian framework).

For choice experiments, the Fisher information matrix depends on the underlying partworth vector, thus the efficiency of a design depends on these unknown parameters. In this context, Huber and Zwerina (1996) show that efficient designs can be achieved when there is some form of prior information on the model parameters. There, the authors identify four guidelines for efficient choice designs (*balance*, *orthogonality*, *minimal overlap* and *utility balance*), and show via numerical experiments that efficiency is improved by focusing on the latter utility balance guideline. Arora and Huber (2001) operationalize this idea by proposing the *aggregate customization* approach to questionnaire design which, roughly speaking, consists on building an efficient (static) design for the average respondent based on the information collected on a pretest. Sándor and Wedel (2001) propose a similar approach in a Bayesian framework, and show that it leads to smaller credibility regions for the parameter estimates. Like in this paper, their focus is on minimizing an approximation to the D-error, while leaving aside traditional design guidelines. That is also the case of Kanninen (2002) which studies D-optimal questionnaires for the asymptotic approximation based on the Fisher information matrix.

Recent work on static designs has focused in the case of consumer heterogeneity, usually in the Bayesian framework. See for example Yu et al. (2009) and Kessels et al. (2009). The latter work, in particular, is among the first (to the best of our knowledge) to suggest the use of individualized adaptive questionnaires. In the same context, Sándor and Wedel (2005) depart from the work above by proposing an *heterogeneous* questionnaire design where different sub-designs are applied to different respondents. More recently, Liu and Tang (2015) propose an heterogeneous questionnaire design with individualized sub-designs, which are shown to outperform extant methods numerically. We refer to this latter work for a review of recent methods for static questionnaire design.

Roughly speaking, the work above, one way or the other, aim at improving efficiency by exploiting the fact that an asymptotic approximation to D-error depends on the partworth vector of each respondent. This, while maintaining the paradigm of static/common questionnaires. Our work, on the other hand, focuses on the design of adaptive questionnaires, which ought to lead to more precise partworth estimates.

Adaptive choice-based Conjoint Analysis. Toubia et al. (2004) are among the first to propose the use of adaptive questionnaires for choice-based tasks. Their approach consists on envisioning the set of possible partworth vectors as a polyhedron, and questions as hyperplanes that cut through said polyhedron. Assuming no response-error, they develop fast polyhedral updates and optimization-based question-selection methods. The resulting method is simple and geometrically intuitive, and, as shown in Hauser and Toubia (2005), it does not seem to suffer from endogeneity bias. Furthermore, the method can be very effective when response-error is low, but its performance can deteriorate when response-error is high.

Toubia et al. (2007) provide a probabilistic interpretation of the polyhedral method and propose a Bayesian extension that incorporates response-error and allows the use of informative priors. In this method, uncertainty over the partworth vector is represented through a mixture of uniform distributions over polyhedra, and response-error modeling favors tractability. Although it preserves the intuitive geometric interpretation of the original method, computation time of question-selection and the Bayesian update scale unfavorably with the number of questions, which limits its application. Bertsimas and O'Hair (2013) present another extension of the polyhedral method, in which response-error is addressed using robust optimization modeling. There, parameter uncertainty is polyhedral, although its representation involves the possibility of reversing a fraction of the respondent's answers. In Section 4 we provide a bayesian interpretation of the geometric methods above, as they are used as benchmark in our numerical experiments.

Based on statistical learning theory, Abernethy et al. (2008) adapt the polyhedral method so that question-selection aims at reducing the uncertainty surrounding parameter estimates when these are computed by minimizing a *regularized network* loss function (Question-selection is implemented heuristically by adapting the utility balance and post-choice symmetry criteria in the original method). Like in the work above, this paper presents an adaptive method for questionnaire design for conjoint analysis somewhat inspired by the polyhedral method.

Our work is also related to Yu et al. (2011), which describe an hybrid static/adaptive full Bayesian model to minimize D-error in the context of choice-based conjoint studies. Using a panel mixed logit model, after an initialization phase, the method computes the posterior distribution of the respondent's partworth vector after each response, which is then used to select the next question so as to minimize the immediate expected D-error, thus effectively resulting on a one-step look-ahead policy. Posterior computation requires multidimensional integration, and question-selection amounts to enumeration (Bliemer and Rose 2010), which renders the approach impractical (see Crabbe et al. (2014) for a similar approach focusing in entropy minimization). The method we propose can be seen as an approximation to a full Bayesian approach in which one strives to maintain practical and theoretical tractability, a clear geometric interpretation and to shed light on the principles guiding optimal question-selection.

Outside the realm of Conjoint Analysis, Toubia et al. (2013) present a non-bayesian methodology for adaptive questionnaire design for estimating preferences. Their focus is on the version of D-efficiency based on the Fisher information matrix. We show how our proposed methods can also work with this alternative definition in Section 5.4.

Conjoint Analysis and Operations Research. Our work applies operations research techniques to address adaptive choice-based conjoint analysis, and as such joins a large list of previous work in marketing. The work above on adaptive questionnaires is a good example, per its use of robust and convex optimization techniques.

Preference estimation is also view as a component of conjoint studies. Like Toubia et al. (2004), Cui and Curry (2005), Evgeniou et al. (2005) and Evgeniou et al. (2007) approach estimation of preferences in conjoint settings as an optimization problem without assuming an underlying probabilistic structure on the data. While Cui and Curry (2005) solve a constrained quadratic problem, Evgeniou et al. (2005) in addition solve nonlinear problems. Focusing on consumer heterogeneity, Evgeniou et al. (2007) adapt the ridge regression method to estimating preferences which amounts to solving a more general convex loss function. The models in this paper are endowed with a probabilistic structure, thus we use the natural Bayesian estimator. However, we do compare our estimator with that in Toubia et al. (2004).

Huang and Luo (2016) propose the use of active machine learning algorithms for adaptive question-selection in the context of preference elicitation for complex products. Their question-selection is guided by the utility balance criterion and uses convex optimization to select the next question. Dzyabura and Hauser (2011) also use machine learning techniques to design adaptive question-selection when respondents use heuristic decision rules. Unlike approaches reviewed so far, their method focuses on minimizing (via enumeration) expected posterior entropy. Like in our paper, their method approximates the posterior distribution so as conduct question-selection without noticeable delay. When explicit enumeration is impractical, question-selection is driven instead by a criterion similar to utility balance.

3. A Bayesian Framework for Optimal Adaptive Conjoint Analysis

3.1. Adaptive questionnaires

We adopt the classic profile representation of products in terms of attributes and levels. In particular, we encode a product profile via a binary vector $x \in \{0,1\}^n$, where a one on the i-th component indicates that the product is endowed with attribute i, and is 0 otherwise. We assume that feasible profiles belong to a set $\mathcal{X} \subseteq \{0,1\}^n$ (without loss of generality, this set may contain an outside alternative).

In the conjoint task, each respondent faces a sequence of questions, which we index by $k \in \{1, \dots, K\}$, where K denotes the total number of questions. Each question consists of a set of product profiles, among which the respondent is asked to select the most preferred one. Here, we assume that each question consists of two profiles, although the analysis extends for the case of more alternatives. We denote by x^k and y^k the first and second profiles presented in question k to a respondent, respectively. Also, we denote by $a_k \in \{1,2\}$ the (random) profiled selected by a respondent on question k.

In this context, a questionnaire is a sequence $\left\{(x^k, y^k), k \le K\right\}$ of questions adapted to the filtration generated by the questions and answers given by the respondent, i.e. (x^k, y^k) is $\mathcal{F}_k$-adapted, where $\mathcal{F}_k := \sigma\left(x^l, y^l, a_l, l < k\right)$, and $\mathcal{F}_1 := \emptyset$.

3.2. Estimation Framework

3.2.1. Respondent Choice We assume that a respondent's (compensatory) preferences are characterized by a vector of *partworths* $\beta \in \mathbb{R}^n$ representing the relative importance she assigns to product attributes. In particular, a respondent assigns a utility U_x to product profile $x \in \{0,1\}^n$, with

$$U_x := \beta \cdot x + \varepsilon_x,$$

where $a \cdot b := \sum_{i=1}^{n} a_i b_i$ denotes the inner product of two vectors $a, b \in \mathbb{R}^n$, and ε_x is a random idiosyncratic shock to utility. We assume that consumer choice follows a logit model, i.e. the shocks to utility are i.i.d. random variables, each following a (standard) Gumbel distribution independent of x, and respondents select the profile with the highest utility among those offered. That is, for $k \in \{1, \ldots, K\}$ we have that a_k is the random variable given by $a_k = 1 + \mathbf{1}\left\{U_{x^k} \le U_{y^k}\right\}$, $k \le K$, where $\mathbf{1}\{\cdot\}$ denotes the indicator function of a set. With this assumption, it is well known that

$$\mathbb{P}(a_k = 1|\beta) = \frac{e^{\beta \cdot x^k}}{e^{\beta \cdot x^k} + e^{\beta \cdot y^k}}. \tag{1}$$

In this context, we say response-error occurs whenever $\beta \cdot x^k > \beta \cdot y^k$ and $U_{x^k} \le U_{y^k}$, or $\beta \cdot x^k \le \beta \cdot y^k$ and $U_{x^k} > U_{y^k}$.

3.2.2. Prior Information, Update and Estimation While the respondents' partworths are unknown upfront, we assume that they are drawn from a prior multivariate normal distribution with a known mean $\mu^0 \in \mathbb{R}^n$ and covariance matrix $\Sigma_0 \in \mathbb{R}^{n \times n}$. Given a respondent's answer to a question this prior distribution is updated using Bayes' rule and (1) as a likelihood function. More precisely, after observing a sequence of questions and answers $\mathcal{F}_k$ we can construct a posterior distribution of a respondent's partworths whose density is given by

$$f(\beta|\mathcal{F}_k) := \frac{\phi(\beta; \mu^0, \Sigma_0)\, L(\mathcal{F}_k|\beta)}{\mathbb{P}(\mathcal{F}_k)},$$

where

$$\phi(\beta; \mu, \Sigma) := \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} e^{-\frac{1}{2}(\beta-\mu)^\top \Sigma^{-1} (\beta-\mu)}$$

denotes the density of a multivariate normal distribution with mean μ and covariance matrix Σ, $\det(\cdot)$ denotes the determinant of a matrix,

$$L(\mathcal{F}_k|\beta) := \prod_{l<k:\, a_l=1} \frac{e^{\beta \cdot x^l}}{e^{\beta \cdot x^l} + e^{\beta \cdot y^l}} \prod_{l<k:\, a_l=2} \frac{e^{\beta \cdot y^l}}{e^{\beta \cdot x^l} + e^{\beta \cdot y^l}}$$

is the likelihood function of $\mathcal{F}_k$ given β and (1), and $\mathbb{P}(\mathcal{F}_k) = \int_{\mathbb{R}^n} \phi(\beta; \mu^0, \Sigma_0)\, L(\mathcal{F}_k|\beta)\, d\beta$. This posterior distribution can be used directly for any stochastic optimization problem that requires information about β or the information can be simplified using a summary statistic. For instance, a natural point-wise estimate of β given the events in $\mathcal{F}_k$ is its expectation under the posterior, i.e.

$$\mu^k := \mathbb{E}(\beta|\mathcal{F}_k) = \int_{\mathbb{R}^n} \beta f(\beta|\mathcal{F}_k)\, d\beta.$$

Similarly, the covariance matrix of the posterior distribution of β can be calculated as $\Sigma_k := \mathbb{E}\left((\beta - \mathbb{E}(\beta|\mathcal{F}_k)) (\beta - \mathbb{E}(\beta|\mathcal{F}_k))^\top \middle| \mathcal{F}_k \right)$. However, both these summary statistics require multidimensional integration, so even approximating them can be computationally expensive.

3.3. Optimal Question Selection Policy

The stated objective of questionnaire design in conjoint experiments is the *precision* of partworth estimates. In the Bayesian context this precision is usually quantified through the expectation of some function of the posterior covariance matrix Σ_k (Yu et al. 2012). In particular, the most widely used efficiency criterion focuses on minimizing the *D-error*, i.e. the expected n-th root of $\det(\Sigma_k)$ when Σ_k is an $n \times n$ matrix. (Huber and Zwerina 1996, Sándor and Wedel 2005).

While we focus on the minimizing the expected n-th root of $\det(\Sigma_k)$ in our experiments, we more broadly consider the d-th root of $\det(\Sigma_k)$, where d is not necessarily equal to n. In particular, we consider the geometric interpretation of the case $d = 2$ which explains one intent of the D-error. This interpretation arises by considering the construction of a confidence set or credibility region of level $\gamma > 0$ for a given distribution (i.e. a region that contains γ of the mass of the distribution). For a multivariate normal distribution with mean μ and covariance matrix Σ, a natural credibility region is given by the ellipsoid

$$\mathcal{E}(\mu, \Sigma, r(\gamma)) := \left\{\beta \in \mathbb{R}^n : (\beta - \mu)^\top \Sigma^{-1}(\beta - \mu) \leq r(\gamma)\right\} \tag{2}$$

where $r(\gamma)$ is the χ^2-value (with n degrees of freedom) located at the γ percentile, so that $\mathbb{P}_{N(\mu,\Sigma)}(\beta \in \mathcal{E}(\mu, \Sigma, r(\gamma))) = \gamma$ (Anderson 1984). The volume of this ellipsoid is a natural dispersion measure and is proportional to $\det(\Sigma)^{1/2}$. While the posterior distribution is no longer multivariate normal, $\mathcal{E}(\mu^k, \Sigma_k, r_k(\gamma))$ for an appropriately chosen $r_k(\cdot)$ should provide a good approximation of a credibility region for relatively small k (e.g. in this setting we expect the posteriors to be nearly multivariate normal or at least unimodal). Hence $\det(\Sigma_k)^{1/2}$ should provide a reasonable dispersion or precision measure. We denote this measure the *volumetric D-error.*

In our Bayesian context, the general D-error $\det(\Sigma_k)^{1/d}$ is a random variable dependent on $\mathcal{F}_k$. Then, an optimal adaptive questionnaire is a non-anticipative strategy π than minimizes the expected final D-error. That is, an optimal strategy π is such that

$$\pi \in \arg\min \left\{\mathbb{E}\left(\det(\Sigma_K)^{1/d}\right) : (x^k, y^k)\, \mathcal{F}_k\text{-measurable for all } k \leq K\right\}. \tag{3}$$

As a sequential decision problem under uncertainty, the formulation above can be casted as a dynamic program (Bellman 1957). Considering the set of questions and answers $\mathcal{F}_k$ observed prior to a question as the state variable, the Bellman (value function) recursion is given by

$$V_k(\mathcal{F}_k) = \min_{x^k, y^k \in \mathcal{X}} \mathbb{E}\left(V_{k+1}(\sigma(\mathcal{F}_k \cup \left\{x^k, y^k, a_k\right\}))\right), \tag{4}$$

with the border condition $V_{K+1}(\mathcal{F}_{K+1}) = \det(\Sigma_{K+1})^{1/d}$. We are interested in $V_1(\emptyset)$. In general, the formulation above is intractable not only because computing posterior distributions require multi-dimensional integration and optimal question-selection can only be resolved via enumeration, but also because the state space scales exponentially with the number of questions and the dimension of the partworth vector, a fact known as the curse of dimensionality of dynamic programming (Powell 2007). A popular heuristic solution in dynamic programming is the so-called one-step look-ahead policy that, in a nutshell, operates in each period as if it was the last period in the horizon. In the context of formulation (4), this translates into treating each question as if it was the last one in the questionnaire. In Section 5 we show how the different geometric methods in the literature can be seen as one-step look-ahead policies, applied over an approximation to the D-error metric.

4. Geometric Benchmark and Prior Information

Next, we use the framework of the previous section to provide an interpretation of the geometric methods reviewed in Section 2, which will serve as benchmark in our numerical experiments.

4.1. Polyhedral Method

Introduced in Toubia et al. (2004), the choice-based Polyhedral method maintains credibility regions in the form of polyhedra, assumes no response-error, and focuses question-selection on the precision of the partworth vector estimate after the credibility region update. In particular, the method begins with a credibility region in the form of a non-empty polyhedron $\mathcal{P}_0 := \{\beta \in \mathbb{R}^n : A\beta \le b\}$ that is believed to contain the true partworth vector. Upon observing the answer to question k, the method updates this credibility region assuming that there is no response-error, i.e. that $\varepsilon_{x^k} = \varepsilon_{y^k}$. Thus, the update preserves only those regions in $\mathcal{P}_0$ that are absolutely "consistent" with a respondent's answer. Letting $\mathcal{P}_k$ denote the polyhedra representing the credibility region after observing the answer to question $k-1$, the update is given by

$$\mathcal{P}_{k+1} := \mathcal{P}_k \cap \begin{cases} \left\{\beta \in \mathbb{R}^n : \beta \cdot x^k \ge \beta \cdot y^k\right\} & \text{if } a_k = 1 \\ \left\{\beta \in \mathbb{R}^n : \beta \cdot x^k \le \beta \cdot y^k\right\} & \text{if } a_k = 2. \end{cases} \tag{5}$$

Note that the representation via a polyhedron is preserved by the update, thus contributing to the simplicity of the method.

A Bayesian interpretation of this approach is that the initial prior distribution on the partworth vector is the uniform distribution on polyhedron $\mathcal{P}_0$ and that the likelihood function is obtained from the choice probability

$$\mathbb{P}\left(a_k = 1 | \beta\right) = \mathbf{1}\left\{\beta \cdot x^k \ge \beta \cdot y^k\right\}.$$

Toubia et al. (2004) propose using analytic center of $\mathcal{P}_{k+1}$ (which can be computed efficiently, see Bertsimas and Tsitsiklis (1997)) as a point-wise estimate of the partworth vector after question k. Standard computation of the analytic center also yields an approximation of the covariance matrix of the uniform distribution over $\mathcal{P}_k$ which can be used to construct an ellipsoidal approximation of $\mathcal{P}_k$ known as Sonnevend's ellipsoid (Sonnevend 1986). With regard to question-selection, the method follows two main guiding principles:

- *Choice-balance*: select question k so that respondents are as indifferent as possible between choices x^k and y^k. Geometrically, this principle aims to minimize the distance between hyperplane $(x^k - y^k) \cdot \beta = 0$ and the center of polyhedron $\mathcal{P}_k$.
- *Post-choice symmetry*: select question k to minimize the maximum variance of any combination of partworths in $\mathcal{P}_{k+1}$. Geometrically, this principle aims for hyperplane $(x^k - y^k) \cdot \beta = 0$ to be as orthogonal as possible to the longest "axis" of polyhedron $\mathcal{P}_k$.

These criteria are heuristically implemented as follows. First, the center and axes of the polyhedron are approximated by the analytic center and axes of the associated Sonnevend's ellipsoid (these can be found efficiently via nonlinear optimization and eigen-decomposition techniques). Then, the intersection of this longest axis (starting from the analytic center) with $\mathcal{P}_k$ yields two partworth vectors β^1 and β^2; x^k and y^k are found solving two independent optimization problems:

$$x^k \in \arg\max\left\{\beta^1 \cdot x : x \in \mathcal{X},\, c \cdot x \leq M\right\} \quad \text{and} \quad y^k \in \arg\max\left\{\beta^2 \cdot x : x \in \mathcal{X},\, c \cdot x \leq M\right\},$$

where $M > 0$ is a (randomly drawn) constant and c is the analytic center of $\mathcal{P}_k$. If x^k happens to be identical to y^k, a new M is resampled and the problems are resolved, repeating until $x^k \neq y^k$.

4.2. Probabilistic Polyhedral Method

Introduced in Toubia et al. (2007), this method generalizes the Polyhedral method by assuming respondents make "mistakes" with constant probability. In its simplest version, the probabilistic polyhedral method follows the Bayesian interpretation of the Polyhedral method using an initial prior distribution that is uniform on polyhedron $\mathcal{P}_0$. However, instead of using the no-error likelihood function obtained from (5), its likelihood is obtained from the choice probability

$$\mathbb{P}(a_k = 1|\beta) = \begin{cases} \alpha & \beta \cdot x^k \geq \beta \cdot y^k \\ 1-\alpha & \beta \cdot x^k < \beta \cdot y^k \end{cases}.$$

where $\alpha \in (0,1)$ is a tuning parameter representing the probability of a "correct" answer.

Although there is no explicit construction of credibility regions, the posterior distribution can be interpreted as a generalization of a polyhedral credibility region as it corresponds to a mixture of uniform distributions over polyhedra. The general version of the probabilistic polyhedral method uses this interpretation to incorporate informative priors that are more complex than simply the uniform distribution on polyhedron $\mathcal{P}_0$. We discuss such informative priors in Section 4.4.

With regard to question-selection, the method follows a direct adaptation of that of the polyhedral method: the analytic center of the mixture of polyhedra is computed as the mixture of the analytic centers of the polyhedra; and the longest axis of a mixture of polyhedra is computed as the vector that maximizes the weighted norm of its projection on the longest axes of the polyhedra in the mixture. While there is a closed-form expression for the longest axis of a mixture, the procedure requires computing the analytic center and longest axis of each polyhedron in said mixture. Unfortunately, the number of polyhedra grows exponentially with the number of questions, which renders their exhaustive enumeration impractical. Because of this, the practical implementation of the question-selection procedure for the probabilistic polyhedral method simply considers a constant number of polyhedra with the largest weights.

4.3. Robust Method

Proposed by Bertsimas and O'Hair (2013), the method adapts the polyhedral method to handle response-error via robust optimization. The method starts with an initial polyhedral credibility region $\mathcal{P}_0$, which is updated by adding only *some* of the linear inequalities associated with a respondent's answers, reversing the sign of the remaining ones to represent the possibility of response-error. Considering a tolerance parameter $\rho < 1$ representing the maximum fraction of responses one is allowed to reverse, the method maintains the credibility region given by

$$\mathcal{P}_{k+1} := \left\{ \beta \in \mathcal{P}_0 : \exists S \subseteq \{1,\ldots,k\} \text{ s.t. } |S| \leq \rho k \quad \wedge \quad \begin{array}{l} \beta \cdot x_l \geq \beta \cdot y_l, \quad \forall\, l \notin S,\ a_l = 1 \text{ or } l \in S,\ a_l = 2 \\ \beta \cdot x_l \leq \beta \cdot y_l, \quad \forall\, l \in S,\ a_l = 1 \text{ or } l \notin S,\ a_l = 2 \end{array} \right\},$$

where S represents the questions in which the respondent's answers are reversed. The resulting credibility region is no longer a polyhedron, but can be represented as the feasible region of a MIP. Bertsimas and O'Hair (2013) obtain a point-wise estimator by combining this MIP representation with the nonlinear optimization problem used to compute the analytic center of a polyhedron. This can be interpreted as a mixed integer nonlinear programming (MINLP) version of the analytic center.

With regard to question-selection, Bertsimas and O'Hair (2013) focus exclusively on the choice-balance criterion. To achieve this, they select the next question by solving (via enumeration)

$$(x^k, y^k) \in \arg\min \left\{ \frac{|(x-y)\cdot c|}{\|x-y\|} : x, y \in \mathcal{X}, \quad x \neq y \right\},$$

where c is a generalization of the analytic center.

4.4. Informative Priors and Ellipsoidal credibility regions

One salient feature of bayesian approaches to question-selection is the possibility of incorporating any prior information on preferences via the prior distribution of the partworth vector. This possibility arises in the context of both static and adaptive questionnaires (see e.g. Sándor and Wedel (2005) Yu et al. (2011)); some work even suggest incorporating information across respondents into adaptive question-selection (see e.g. Dzyabura and Hauser (2011), Huang and Luo (2016)).

In this regard, one way the probabilistic polyhedral method improves upon the polyhedral method is by incorporating prior information. In particular, Toubia et al. (2007) show how a mixture of polyhedra can be used to approximate a multivariate normal prior over the partworth vector β at the cost of using mixtures of uniform distributions on polyhedra. An alternative is to use the LOG-Het approach proposed by Evgeniou et al. (2007) to incorporate multivariate priors through a regularization in the analytic center estimation procedure. This approach was incorporated into the question-selection procedure for standard multivariate normal priors by Abernethy et al. (2008) and can be directly extended to general multivariate priors. However, it requires selecting a penalty parameter that can be hard to tune in the context of individualized partworth estimate procedures. One might avoid selecting such a parameter by using Bayesian methods such as the one described in Section 3. However, Bayesian methods replace the parameter selection with potentially costly multidimensional integration.

Fortunately, one can incorporate multivariate normal priors to all geometric methods by considering ellipsoidal credibility regions such as (2). To achieve this, instead of starting from an initial polyhedral credibility region $\mathcal{P}_0$ we may instead start from an initial ellipsoidal credibility region given by

$$\mathcal{P}_0 := \mathcal{E}\left(\mu^0, \Sigma_0, r(\rho)\right) \tag{6}$$

for some confidence level ρ, where $\mu^0 \in \mathbb{R}^n$ and $\Sigma_0 \in \mathbb{R}^{n\times n}$ are the mean and covariance matrix of the multivariate normal prior. This modification can be easily incorporated into all the geometric methods described in the previous sections. For instance, the analytic center and Sonnevend's ellipsoid have straightforward extensions to the intersection of an ellipsoid and polyhedron (Boyd and Vandenberghe 2004). Furthermore, incorporating such ellipsoidal priors only incurs in a marginal computational cost when using state-of-the-art optimization techniques. In addition, it yields a way to incorporate informative priors with a simple geometric interpretation. For these reasons from now on we assume that all geometric methods (benchmark described so far and new methods described in the following sections) begin from the initial ellipsoidal credibility region (6).

One of the advantages of having ellipsoidal credibility regions is that getting a point-wise estimate and a dispersion measure is straightforward.[2] Unfortunately, this simplicity disappears if we update the region with any of the methods described so far. This also applies to the proper Bayes update of a normal prior distribution of the partworth vector: the posterior distribution is not elliptical, and thus credibility regions are not ellipsoids. However, for the Polyhedral method we could exploit the fact that we can efficiently compute the minimum volume ellipsoid containing the intersection of an ellipsoid and a half-space defined by a linear inequality (see, e.g. Grötschel et al. (2011)). Hence, we could simply replace such ellipsoid/half-space intersection obtained after the first question with the corresponding minimum volume ellipsoid. Repeating this for each question would allow us to maintain an ellipsoidal credibility region. In preliminary experiments, this approach provided a tangible improvement to the polyhedral method and a similar approach has been used in the context of pricing in Cohen et al. (2016). However, an undesirable artifact of the minimum volume ellipsoid is an increment in variance in directions orthogonal to the question defining the half-space. We can use some linear algebra to correct this variance increment, but our preliminary experiments showed that both these minimum ellipsoid approaches are dominated by the Bayesian-based ellipsoidal approximation of the posterior credibility region we describe in the following section.

5. The Approximate Bayesian Ellipsoidal Method

The attractiveness of the polyhedral method follows mainly from: i) its simplicity and clear geometric intuition; and ii) its suitability for incorporating optimization techniques into question-selection. Unfortunately, these features mostly disappear in the extensions discussed in the previous section. For example, the clear interpretation of $\mathcal{P}_k$ as a credibility region is lost in the probabilistic and robust adaptations of the method. Also, the steps required to either perform a region/distribution update or to select a question get more convoluted (cf. the notion of the longest axis of a mixture of polyhedra and the proposed generalization of the analytic center in the robust approach).

Next, we propose a method that approximates the optimal adaptive questionnaire while preserving the positive features of the polyhedral method mentioned above. In addition, the method maintains the treatment of response-errors of the Bayesian approach. At the core of the method are the following approximations.

- Following all benchmark methods from Section 4, we focus on a one-step look-ahead approximate dynamic programming policy that assumes the next question is the last.
- We use a moment matching approximate Bayesian approach to approximate all posterior distributions with a multivariate normal distribution in a computationally efficient way.
- We combine approximate Bayesian and MIP techniques to approximate the optimal question-selection problem as a MIP.

The first approximation above allows us to approximate the dynamic programming formulation (4) with $V_k(\mathcal{F}_k) = \min_{x^k,y^k \in \mathcal{X}} \mathbb{E}\left(\det\left(\Sigma_{k+1}\right)^{1/d}\right)$; the second approximation allows us to replace the state variable $\mathcal{F}_k$ with the mean μ^k and covariance Σ_k of the approximate normal posterior; the third approximation replaces $\mathbb{E}\left(\det\left(\Sigma_{k+1}\right)^{1/d}\right)$ with a MIP-based approximation $\breve{g}$ that depends on (x^k, y^k), μ^k and Σ_k. With this, the approximate dynamic program becomes

$$V_k(\mu_k, \Sigma_k) = \min_{x^k,y^k \in \mathcal{X}} \breve{g}\left((x^k, y^k), \mu^k, \Sigma_k\right).$$

In the sequel, we show that the minimization above, as well as the computation of the normal posterior approximation, can be carried out almost instantaneously for practical instances of the problem, and thus can be implemented in a real-time environment. Moreover, we show that by envisioning the posterior update in terms of the underlying ellipsoidal credibility regions, the method admits a simple and clear geometrical interpretation.

5.1. Efficient Approximate Posterior Updates

As argued in Section 3.3, a multivariate normal distribution should provide a reasonable approximation to the posterior distribution of the partworth vector (after a single question) when its prior is also multivariate normal. This is precisely the premise of moment-matching approximation approaches (e.g. (Gelman et al. 2013, Section 13.8)). The main difficulty for constructing the approximate normal posterior is the numerical computation of the posterior's mean and covariance matrix, which often involves approximating a series of n-dimensional integrals. Fortunately, moment matching approximations for Bayesian logistic regression problems like the one we face can be efficiently computed by approximating a series of one-dimensional integrals (e.g. (Gelman et al. 2013, page 340)). The following proposition, whose self-contained proof is included in Appendix A, formalizes this result.

Proposition 1. *For two profiles $x, y \in \mathcal{X}$ such that $x \neq y$, let $\Omega := \{\beta \in \mathbb{R}^n : U_x \geq U_y\}$ and suppose that $\beta \sim N(\mu, \Sigma)$. Furthermore, let*

- $\Sigma^{1/2}$ *be the square root or Cholesky decomposition of* Σ *such that* $\Sigma = \Sigma^{1/2}\left(\Sigma^{1/2}\right)^\top$,
- $\mu_{x,y} := (x-y)\cdot\mu$ *and* $\sigma_{x,y} := \left\|\left(\Sigma^{1/2}\right)^\top (x-y)\right\|$ *be the mean and variance of question* $(x-y)\cdot\beta$,
- $W = [w^1, \ldots, w^n] \in \mathbb{R}^{n\times n}$ *where* $\{w^i\}_{i=1}^n$ *is an orthonormal basis such that* $w^1 = \left(\Sigma^{1/2}\right)^\top (x-y)/\sigma_{x,y}$, *and*
- Z *is a random variable with density*

$$f_Z(z) := \frac{\left(1+e^{-\mu_{x,y}-\sigma_{x,y}z}\right)^{-1}\phi(z)}{\int_{\mathbb{R}}\left(1+e^{-\mu_{x,y}-\sigma_{x,y}z}\right)^{-1}\phi(z)dz},$$

where $\phi(\cdot)$ denotes the standard normal density.

Then,

$$\mathbb{E}(\beta|\Omega) = \mu + \left(\Sigma^{1/2}W\right)_1 \mathbb{E}(Z) \quad \text{and} \quad Cov(\beta|\Omega) = \left(\Sigma^{1/2}W\right)\begin{pmatrix}\mathrm{Var}(Z) & 0\\ 0 & I\end{pmatrix}\left(\Sigma^{1/2}W\right)^\top, \tag{7}$$

where $\left(\Sigma^{1/2}W\right)_1$ denotes the first column of matrix $\Sigma^{1/2}W$.

Proposition 1 allows for the fast computation of approximate posteriors, thus keeping the method computationally tractable and practical. Indeed, the moments of Z above can be efficiently computed through one-dimensional numerical integration or Montecarlo simulation (in our implementation, we chose the former). In terms of the geometry of the method, as mentioned before, one can envision the approximate posterior update in terms of the credibility regions of the prior and posterior distributions: the ellipsoidal credibility region is updated as follows

$$\mathcal{E}\left(\mu^k, \Sigma_k, r(\rho)\right) \to \mathcal{E}\left(\mu^{k+1}, \Sigma_{k+1}, r(\rho)\right),$$

where the dependence between approximate prior and posterior parameters, for a given question and answer, is that specified by (7).

Next, we show that Proposition 1 allows us not only to work efficiently with the posterior distribution, but also to write a simple optimization problem to conduct question-selection.

5.2. Efficient Approximate Optimal Question Selection

In this section, we use Proposition 1 to write D-error minimization via MIP. The next result shows that, for a given question, evaluation of D-error can be conducted via one-dimensional integration. In addition, it sheds light into the importance of choice-balance and post-choice symmetry and more importantly, yields the precise trade-off between them to minimize D-error.

PROPOSITION 2. *For two profiles $x, y \in \mathcal{X}$ such that $x \neq y$, let $a := 1 + \mathbf{1}\{U_x \leq U_y\}$ and suppose that $\beta \sim N(\mu, \Sigma)$. Furthermore, let*

- *$\mu_{x,y}$ and $\sigma_{x,y}$ be the mean and variance of question $(x-y)\cdot\beta$ as defined in Proposition 1,*
- *$p(m,v) := \int_{\mathbb{R}} \left(1+e^{-m-vz}\right)^{-1}\phi(z)dz$, and $Z(m,v)$ be a random variable with density*

$$f_{Z(m,v)}(z) := \frac{\left(1+e^{-m-vz}\right)^{-1}\phi(z)}{p(m,v)},$$

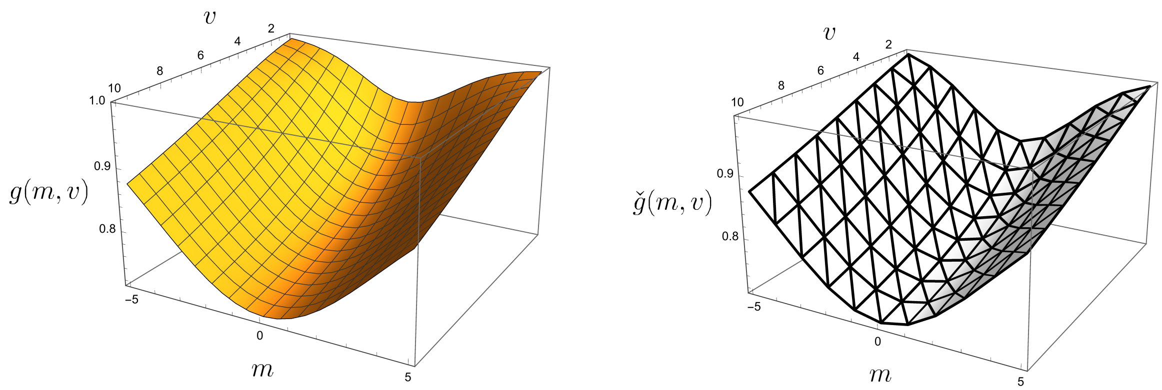
where $\phi(\cdot)$ denotes the standard normal density and $m, v \in \mathbb{R}$ with $v > 0$.

Then, for any integer $d \geq 2$ we have

$$\mathbb{E}_a\left(\det\left(Cov\left(\beta|a\right)\right)^{1/d}\right) := \det\left(\Sigma\right)^{1/d} g\left(\mu_{x,y}, \sigma_{x,y}\right),$$

where $g(m,v) := p(m,v)\,\mathrm{Var}\left(Z(m,v)\right)^{1/d} + p(-m,v)\,\mathrm{Var}\left(Z(-m,v)\right)^{1/d}$.

Proposition 2 shows that the expected D-error is proportional to a two-variable function g of the mean and variance of question $(x-y)\cdot\beta$ under $\beta \sim N(\mu,\Sigma)$. Such a two-variable function is depicted in Figure 1(a) for the volumetric D-error which corresponds to $d = 2$. From Figure 1(a) we see that the D-error is smaller the closer the question mean $(x-y)\cdot\mu$ is to zero. In terms of the geometry of the method, the closer the question mean is to zero, the closer the question cuts through the center of the prior ellipsoid. Hence, the D-error is smaller the closer we are to satisfying choice-balance. Similarly, we see that the D-error decreases as the question variance $\left\|\left(\Sigma^{1/2}\right)^\top(x-y)\right\| = (x-y)^\top\Sigma(x-y)$ increases. If the question cuts through the center of the



(a) Exact function. (b) Piecewise linear approximation.

Figure 1 Two variable function relating question mean and variance with D-error.

prior ellipsoid, the question variance is maximized precisely when the question is orthogonal to the largest axis of the prior ellipsoid. Hence, the D-error is smaller the closer we are to satisfying post-choice symmetry. As we study further in Section 5.2, this suggests that the two principles used for question-selection in the polyhedral method are precisely those that, if properly balanced, yield the smallest expected D-error. Furthermore, Proposition 1 gives us a precise way to optimize D-error over a restricted set of questions through

$$\min\left\{g\left((x-y)\cdot\mu, \left\|\left(\Sigma^{1/2}\right)^{\top}(x-y)\right\|\right) : x, y \in \mathcal{X}, \quad x \neq y\right\}. \tag{8}$$

As written, (8) is still impractical as it is a non-convex MIP whose objective function can only be evaluated numerically. However, numerical computation is only necessary for the evaluation of $g(m, v)$ at different points (m, v) and such a computation only requires one-dimensional numerical integration or Montecarlo simulation. Furthermore, g is only a two-variable function, so by evaluating it in a grid for the possible values of m and v we can easily construct a piecewise linear function $\breve{g}$ that closely approximates g. One such function is depicted in Figure 1(b). We can use any such piecewise linear approximation and standard MIP techniques to construct the following formulation of (8).

$$\min \quad \breve{g}\left(\sum_{i=1}^{n}(x_i-y_i)\mu_i, \sum_{i=1}^{n}\sum_{j=1}^{n}\left(X_{i,j}+Y_{i,j}-W_{i,j}-W_{j,i}\right)\Sigma_{i,j}\right) \tag{9a}$$

$s.t.$

$$X_{i,j} \leq x_i, \quad X_{i,j} \leq x_j, \quad X_{i,j} \geq x_i + x_j - 1 \quad \forall i, j \in \{1, \ldots, n\} \tag{9b}$$

$$Y_{i,j} \leq y_i, \quad Y_{i,j} \leq y_j, \quad Y_{i,j} \geq y_i + y_j - 1 \quad \forall i, j \in \{1, \ldots, n\} \tag{9c}$$

$$W_{i,j} \leq x_i, \quad W_{i,j} \leq y_j, \quad W_{i,j} \geq x_i + y_j - 1 \quad \forall i, j \in \{1, \ldots, n\} \tag{9d}$$

$$\sum\nolimits_{i=1}^{n} X_{i,i} + Y_{i,i} - 2W_{i,j} \geq 1 \tag{9e}$$

$$x, y \in \mathcal{X} \subseteq \{0,1\}^n \tag{9f}$$

$$X, Y, W \in \{0,1\}^{n\times n}. \tag{9g}$$

Constraints (9b)–(9d) and (9g) enforce $X_{i,j} = x_i x_j$, $Y_{i,j} = y_i y_j$ and $W_{i,j} = x_i y_j$ for all i, j and constraint (9e) enforces $x \neq y$. If $\mathcal{X}$ can be described by linear inequalities (plus the integrality of x, y), then the only non-linearity of (9) is the piecewise linear objective. However, such objective can be modeled exactly as MIP using standard techniques (Huchette and Vielma 2017, Vielma et al. 2010). Applying such techniques yields a linear MIP that can be effectively solved with state-of-the-art solvers. This allows for a near-optimal selection of a question.

5.3. Method Summary

The proposed method approximates a one-step look-ahead optimal policy for the dynamic programming formulation of the optimal adaptive questionnaire (3). For this, after observing each of the respondent's answers, we use (7) to approximate the partworths' posterior distribution by a multivariate normal with the same mean and covariance matrix of the posterior, which one can compute efficiently as per Proposition 1. With this approximation, question-selection focuses in minimizing the expected D-error: this step can be can be further approximated by solving the MIP formulation (9). We formalize the proposed method next.

Definition 1 (Ellipsoidal Method). Let $\mu^0 \in \mathbb{R}^n$ and $\Sigma_0 \in \mathbb{R}^{n\times n}$ be such that the initial prior distribution of the partworth vector is $\beta \sim N(\mu^0, \Sigma_0)$. The policy is given by the following recursion for all $k \in \{1, \ldots, K\}$.

- (μ^k, Σ_k) are calculated through equation (7) of Proposition 1 for $\mu = \mu^{k-1}$, $\Sigma = \Sigma_{k-1}$, $x = x^{k-1}$ and $y = y^{k-1}$, and
- (x^k, y^k) are an optimal solution of (9) for $\mu = \mu^{k-1}$, $\Sigma = \Sigma_{k-1}$, $x = x^{k-1}$ and $y = y^{k-1}$.

5.4. Other Variance Measures

The key property given by Proposition 2 is that the expected posterior D-error only depends on the question profiles x, y through $\mu_{x,y}$ and $\sigma_{x,y}$. From the proof of Proposition 2 we can see that this also holds for more general variance measures of the form $\mathbb{E}_a\left(\psi\left(\det\left(Cov\left(\beta|a\right)\right)\right)\right)$ for $\psi : \mathbb{R} \to \mathbb{R}$. In particular, it holds for the expected Shannon information or entropy $\mathbb{E}_a\left(\log\left(\det\left(Cov\left(\beta|a\right)\right)\right)\right)$. This does not hold for arbitrary functions of $Cov\left(\beta|a\right)$ and in particular fails for expected A-, C-, and E-errors.[3] However, it does hold for other measures of posterior variance that are not based on the posterior covariance matrix. In particular, Proposition 3 below shows it does hold for some measures based on the generalized Fisher information matrix given by the negative of the expected Hessian of the log posterior likelihood (e.g. (Yu et al. 2012, Section 4.2)). In our context this generalized Fisher information matrix is given by

$$H(\alpha) := -\mathbb{E}_a\left(\text{Hess}\left(\ln f(\beta|a)\right)|_{\beta=\alpha}\right) = \left(\left(1+e^{-z\cdot\alpha}\right)^{-1}\left(1+e^{z\cdot\alpha}\right)^{-1} z\,z^\top + \Sigma^{-1}\right), \tag{10}$$

where $z := (x-y)$ and $z\,z^\top$ denotes the outer product of z with itself. The inverse of $H(\alpha)$ is often used to approximate the posterior covariance matrix and, in fact, it is common to define D-error based on $\det(H(\alpha))^{-1/d}$. Dependency of α can be eliminated by taking expectation with respect to the prior distribution, but, in principle, this requires expensive multidimensional integration, so a computationally effective alternative is to evaluate $H(\alpha)$ with α equal to the posterior mean (Toubia et al. 2013). Fortunately, Proposition 3 shows that the expectation can be achieved via univariate integration and both Fisher-based versions of D-error depend on the question only through $\mu_{x,y}$ and $\sigma_{x,y}$. A similar result also holds for Fisher-based versions of entropy and for the minimum volume ellipsoid-based measures of variance used in our preliminary experiments and in Cohen et al. (2016) (see the comments at the end of Section 4.4).

Proposition 3. *For two profiles $x, y \in \mathcal{X}$ such that $x \neq y$, let $a := 1 + \mathbf{1}\{U_x \leq U_y\}$ and suppose that $\beta \sim N(\mu, \Sigma)$. Furthermore, let $\mu_{x,y}$ and $\sigma_{x,y}$ be the mean and and variance of question $(x-y)\cdot\beta$ as defined in Proposition 1, and $\psi : \mathbb{R} \to \mathbb{R}$ be an arbitrary function. There exists a function $g_\psi : \mathbb{R}^2 \to \mathbb{R}$, which can be computed through univariate integration and for which*

$$\mathbb{E}_a\left(\psi\left(\det\left(Cov\left(\beta|a\right)\right)\right)\right) := \psi\left(\det(\Sigma)\ \ g_\psi\left(\mu_{x,y}, \sigma_{x,y}\right)\right).$$

In addition, if $H(\alpha)$ is the generalized Fisher information matrix defined in (10) *then*

$$\psi\left(\det\left(H(\mu)\right)^{-1}\right) = \psi\left(\det(\Sigma)\left(1 + \frac{\sigma_{x,y}^2}{2+2\cosh\left(\mu_{x,y}\right)}\right)^{-1}\right) \tag{11}$$

and

$$\mathbb{E}_{\alpha\sim N(\mu,\Sigma)}\left(\psi\left(\det\left(H(\alpha)\right)^{-1}\right)\right) = \int_{\mathbb{R}} \psi\left(\det(\Sigma)\left(1 + \frac{\sigma_{x,y}^2}{2+2\cosh\left(\mu_{x,y}+\sigma_{x,y}z\right)}\right)^{-1}\right)\phi(z)dz. \tag{12}$$

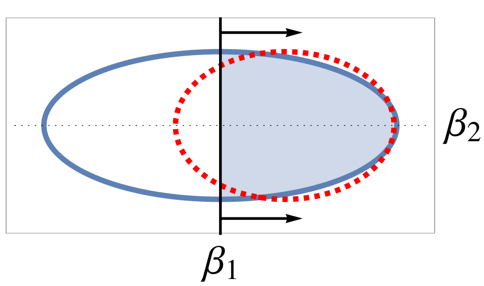
Finally, let $\mathcal{P}(a)$ denote the set of partworths that are contained in the confidence ellipsoid $\mathcal{E}(\mu, \Sigma, r)$ that are also consistent with answer a, absent idiosyncratic shocks to utility (i.e. $\mathcal{P}(a) := \{\beta \in \mathcal{E}(\mu, \Sigma, r) : (2a-3)(x-y)\cdot\beta \geq 0\}$). If $\mathcal{M}(a)$ is the minimum volume ellipsoid containing $\mathcal{P}(a)$ and $\mathcal{P}(1)\cap\mathcal{P}(2) \neq \emptyset$, then there exists $g : \mathbb{R}^3 \to \mathbb{R}$ such that

$$\mathbb{E}_a\left(\text{vol}\left(\mathcal{M}(a)\right)\right) = \det(\Sigma)^{1/2}\, g\left(\mu_{x,y}, \sigma_{x,y}, n\right).$$

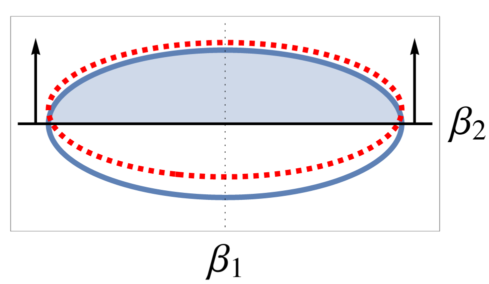
All the measures considered in Proposition 3 allow for a formulation of a MIP akin to (9). In addition, constructing the formulation for the measures based on evaluating the generalized fisher information matrix on μ as done in (Toubia et al. 2013) does not require numerical integration. Thus, it provides an interesting approximation to the exact expected posterior D-error. In Appendix B.3 we show that a variant of the ellipsoidal method based on this approach performs nearly identically to the base method in our set of experiments.

5.5. Further Geometric Interpretation of Optimal Question Selection

As previously mentioned, the relationship between D-error, choice-balance and post-choice symmetry give some interesting insights into these last two criteria. Consider Figure 2, where the blue (solid) and red (dotted) ellipsoids represent prior and posterior credibility regions, respectively. This figure illustrates how the choice-balance criterion of cutting through the center of the ellipsoid is advantageous as it minimizes the expected volume of the resultant half-ellipsoid by inducing two half-ellipsoids with the same volume.[4] In contrast, the advantage of post-choice symmetry is less evident as the expected volume of resultant half-ellipsoid for both questions in Figure 2 is identical as they both cut through the center of the prior ellipsoid. However, the advantage becomes clear by noting that the volume of the posterior ellipsoid for the question that cuts perpendicular to the longest axis of the prior ellipsoid (Figure 2(a)) is about 20% smaller than the one for the question that cuts perpendicular to the shortest axis.[5]



(a) Question perpendicular to the longest axis.



(b) Question perpendicular to the shortest axis.

Figure 2 Ellipsoidal credibility regions for prior (depicted by solid blue ellipsoid) and approximate normal posterior (dashed red ellipsoids) for different questions. The axes of the prior ellipsoid are depicted by the dotted grey line.

Ideally, one would like to find questions that go through the center of the prior ellipsoid (best possible choice-balance) and are orthogonal to the largest axis of the prior ellipsoid (best possible post-choice symmetry). Constraints on the questions may prevent simultaneously satisfying both criteria. However, Figure 1(a) illustrates how, subject to one of the criteria being fixed, improving the other criteria is precisely what is needed to minimize D-error. This can be further confirmed by examining the Fisher-based approximation of D-error from (11) in Proposition 3. This approximation is given by

$$F(\mu_{x,y},\sigma_{x,y},\Sigma) := \det(\Sigma)^{1/d}\left(1+\frac{\sigma_{x,y}^2}{2+2\cosh(\mu_{x,y})}\right)^{-1/d},$$

for which we can easily check that $\frac{\partial F(m,v,\Sigma)}{\partial v} \le 0$ and $m\frac{\partial F(m,v,\Sigma)}{\partial m} \ge 0$ for all m,Σ and $v \ge 0$.

6. Numerical Experiments

In this section we show how the policy from Definition 1 can be implemented in real-time, and that it significantly outperforms the benchmark from Section 4. In this regard, we compare the ellipsoidal method from Definition 1 to all three geometric methods described in Section 4.

6.1. Settings and Implementation details

Experimental setup. To compare the effectiveness of the different methods we conduct Monte-carlo simulation experiments similar to those in Toubia et al. (2004, 2007). That is, we consider a multivariate normal prior distribution of the partworth vector with mean μ^0 and covariance matrix Σ_0 under four regimes that vary response accuracy and population heterogeneity. To model response accuracy we consider $\mu_j^0 = 0.5$ in the low accuracy regime and $\mu_j^0 = 1.5$ in the high accuracy regime, for $j = 1,\ldots,n$. For heterogeneity we consider a diagonal Σ_0 with $(\Sigma_0)_{j,j} = 0.5 \times \mu_j^0$ in the low heterogeneity regime and with $(\Sigma_0)_{j,j} = 2.0 \times \mu_j^0$ in the high heterogeneity regime, for $j = 1,\ldots,n$. For each of the four possible combinations of accuracy/heterogeneity regimes we sample 100 partworths from the corresponding prior distribution, which we consider the "true" partworths of 100 individuals. Then, for each individual we run a simulation where we ask two-product choice-based questions that the individuals respond using their true partworth vector and the random utility model described in Section 3.2.1. For this we consider $n = 12$.

Real-time Implementation. All methods were implemented in the Julia programming language (Bezanson et al. 2017). Computation of the moments for the approximate posterior update in equation (7) was conducted via one-dimensional numerical integration using adaptive Gauss-Kronrod quadrature (*QuadGK.jl* 2017): associated computation times were negligible in all instances.

Optimization problems for question-selection for all methods were modeled using JuMP (Dunning et al. 2017, Lubin and Dunning 2015) and solved using CPLEX 12.6.3 (IBM ILOG 2015) on a Core i7-3770 3.40Ghz desktop computer (Late 2012 iMac). For the case of the proposed method, computation of the piecewise linear approximation $\check{g}$ in (9) was automated using available Julia packages (Huchette 2017, Huchette and Vielma 2017). In all cases $\check{g}$ was constructed using an 8×8

grid. While solution time for problem (9) was capped at one second, this limit was never reached in the low accuracy regimes and it was rarely reached for the high accuracy regimes. The overall average question-selection processing time for the low and high accuracy regimes were 0.19 and 0.29 seconds.

Open-source implementations of all methods in the paper are available at (Sauré and Vielma 2017).

Benchmark. We compare the ellipsoidal method with the three benchmark methods described in Section 4. As described in Section 4.4 an ellipsoidal initial prior credibility can be trivially incorporated to all three benchmark methods, so we let all four methods start with the same initial ellipsoid. Hence, the main differences with the methods are (1) the way they estimate the partworths, (2) the way they model error to update the credibility region, and (3) the way they adaptively select questions. Table 1 summarizes these characteristics for all four methods.

Method	Estimator	Error Modeling	Question Selection
Ellipsoidal	Approx. Bayesian	Approx. Bayesian	Approx. D-error + MIP
Polyhedral	Analytic center	No error	Heuristic
Prob. Polyhedral	Mixture of analytic centers	Fixed error probability	Heuristic
Robust	MINLP analytic center	Fixed fraction of errors	Heuristic

Table 1 Method Characteristics

6.2. Quality of Method's Estimator

Let β^i be the true partworth vector of individual $i \in \{1, \ldots, 100\}$. We begin by measuring the quality of each method's point-wise estimator after question $k \in \{0, 1, \ldots, K\}$ for individual i, which we denote by $\hat{\beta}^{k,i}$. For all methods and individuals i we have that $\hat{\beta}^{0,i}$ is equal to μ^0, the mean of the posterior distribution. For $k > 1$ the estimator varies with the method as described in Sections 4 and 5, and summarized in Table 1.

Table 2 presents summaries of various performance metrics for the estimators right after the fourth, eight and sixteenth question ($k \in \{4, 8, 16\}$). Following Toubia et al. (2007) we report the root mean squared error (RMSE) of the estimator after normalizing it and the true partworth vector so that the sum of the absolute values of the coefficients for all features is equal to the number of parameters. In addition, following Huber et al. (1993), we consider two out-of-sample metrics. The first one is the individual hit rates on 1000 holdout questions. The second is the mean absolute error (MAE) in predicting marketshares for the same 1000 holdout questions. Unlike the Bayesian estimators considered later, the different methods do not provide a uniform notion of covariance matrix for their estimators. For this reason we evaluate the Fisher-based approximation of D-error from (11) with μ equal to the corresponding estimator as a proxy for the true D-Error (evaluation at the estimator yielded a better approximation than evaluating at the prior mean). Table 2 presents averages for all these metrics. In the table, a bold underlined value indicates that the associated method is significantly better ($\alpha = 0.05$) than all other methods for the corresponding metric.

	Fisher D-Error			RMSE			Hit-Rate			Marketshare MAE		
# Questions	4	8	16	4	8	16	4	8	16	4	8	16
Low A, Low H												
Ellipsoidal	**<u>0.21</u>**	**<u>0.18</u>**	**<u>0.14</u>**	**<u>0.84</u>**	**<u>0.76</u>**	**<u>0.67</u>**	**<u>0.69</u>**	**<u>0.71</u>**	**<u>0.72</u>**	0.05	0.05	0.04
Polyhedral	0.24	0.23	0.21	1.08	1.02	0.99	0.64	0.65	0.67	0.04	0.04	0.04
Prob. Poly.	0.24	0.23	0.21	0.91	0.97	1.00	0.68	0.67	0.67	0.05	0.04	0.04
Robust	0.24	0.24	0.22	1.06	1.14	1.10	0.65	0.64	0.65	0.06	0.05	0.04
High A, Low H												
Ellipsoidal	**<u>0.61</u>**	**<u>0.49</u>**	**<u>0.33</u>**	**<u>0.50</u>**	**<u>0.45</u>**	**<u>0.37</u>**	**<u>0.81</u>**	**<u>0.83</u>**	**<u>0.84</u>**	0.06	0.05	0.03
Polyhedral	0.73	0.68	0.56	0.64	0.63	0.60	0.78	0.78	0.79	**<u>0.03</u>**	0.03	0.03
Prob. Poly.	0.71	0.66	0.55	0.58	0.57	0.57	0.79	0.80	0.80	0.04	0.03	0.03
Robust	0.72	0.67	0.59	0.66	0.71	0.66	0.78	0.76	0.78	0.05	0.03	0.03
Low A, High H												
Ellipsoidal	**<u>0.81</u>**	**<u>0.66</u>**	**<u>0.42</u>**	**<u>1.10</u>**	**<u>0.98</u>**	**<u>0.76</u>**	**<u>0.68</u>**	**<u>0.71</u>**	**<u>0.76</u>**	0.07	0.05	0.04
Polyhedral	0.97	0.89	0.68	1.31	1.16	0.99	0.62	0.66	0.70	**<u>0.05</u>**	**<u>0.04</u>**	0.04
Prob. Poly.	0.90	0.81	0.64	1.19	1.10	0.97	0.64	0.67	0.71	0.07	0.05	0.04
Robust	0.95	0.90	0.79	1.38	1.40	1.31	0.60	0.59	0.62	0.06	0.05	0.04
High A, High H												
Ellipsoidal	**<u>2.71</u>**	**<u>2.44</u>**	**<u>1.38</u>**	0.89	**<u>0.77</u>**	**<u>0.55</u>**	0.75	**<u>0.78</u>**	**<u>0.84</u>**	0.10	0.07	0.04
Polyhedral	2.95	2.77	2.05	0.97	0.86	0.68	0.72	0.75	0.80	**<u>0.05</u>**	**<u>0.04</u>**	**<u>0.03</u>**
Prob. Poly.	2.85	2.57	1.92	0.93	0.82	0.65	0.73	0.76	0.81	0.08	0.06	0.04
Robust	2.91	2.70	2.22	1.12	1.13	0.98	0.68	0.69	0.73	0.06	0.05	0.05

Table 2 Summary of Quality Measures for Method's Estimator

We can see that the ellipsoidal method provides a rather consistent advantage over the benchmarks for all metric except for marketshare prediction. Such an inconsistency has been documented in the literature[6] and it is mitigated when one considers a common estimator for all methods (see Tables 3 and 4 and the associated comments in the next section.)

To get a more detailed view of the temporal evolution of the estimators as more questions are answered we present plots of the metrics as a function of the number of questions answered in Figure 3. To complement the RMSE we use a more geometric and individualized measure of the estimators' quality. This measure calculates the geometric or euclidean distance between the estimator $\hat{\beta}^{k,i}$ and the true partworth vector β^i. Considering that only the relative weights between the components of the partworth vector are important, we again normalize the vectors as for RMSE. However, following the geometric nature of this measure we normalize them so that their euclidean norm is equal to one (For RMSE the normalization was for the $l1$ or Manhattan norm to be equal to the length of β). However, even with this normalization, the distance between β^i and the initial estimator μ^0 may vary greatly between individuals. For this reason, we further normalize by this initial distance. That, is for each method we calculate

$$d_{i,k} = \left\| \frac{\hat{\beta}^{k,i}}{\|\hat{\beta}^k\|} - \frac{\beta^i}{\|\beta^i\|} \right\| \Big/ \left\| \frac{\mu^0}{\|\mu^0\|} - \frac{\beta^i}{\|\beta^i\|} \right\| \quad \forall k \in \{0, 1, \ldots, 16\}, i \in \{1, \ldots, 100\}.$$

For all methods and individuals i we have that $d_{i,0} = 1$. Hence, $d_{i,k} < 1$ means that the questions have improved the original estimate and $d_{i,k} > 1$ means that they have worsened it. Figure 3 presents all metrics for the low accuracy and high heterogeneity regime. For D-Error we plot the median and a confidence interval between the first and third quartiles. For the other metrics we shrink the interval to that between the seventh and thirteenth ventiles to improve clarity in the figure (the quartile intervals overlapped too much for these metrics). Figure 3 again shows that the

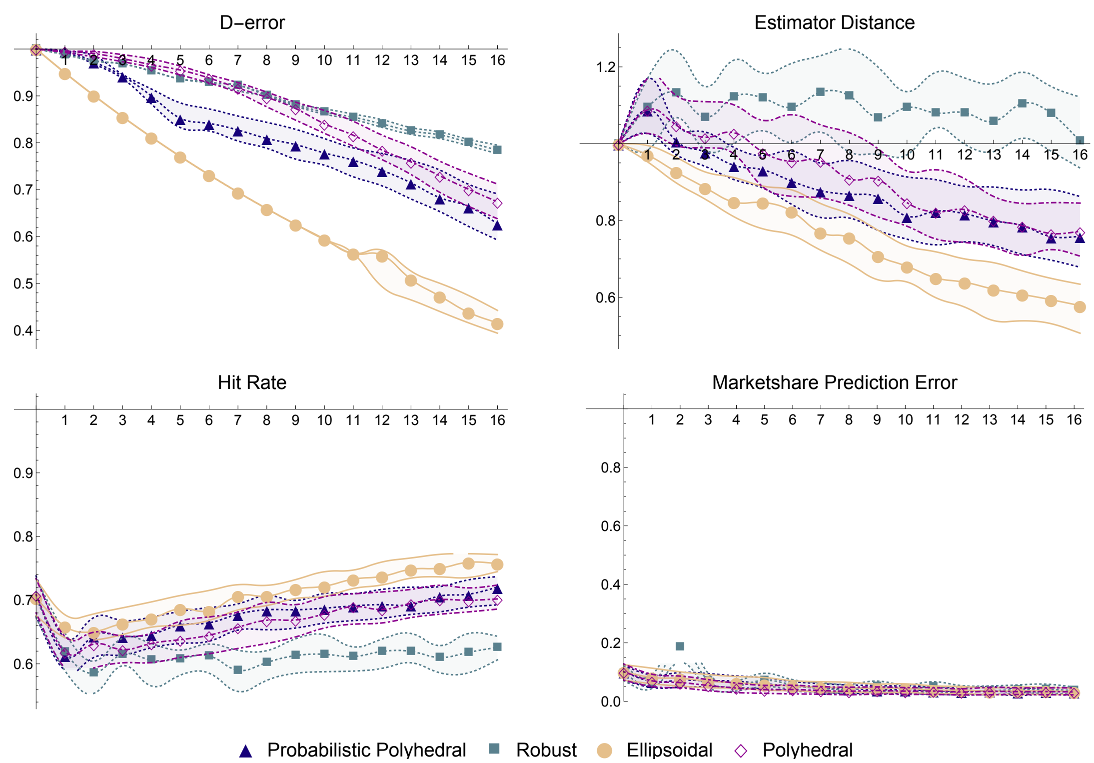


Figure 3 Temporal Evolution of Quality Measures for Method's Estimator in the Low Accuracy and High Heterogeneity Regime

ellipsoidal method provides a significant advantage on the quality of the point-wise estimator and on hit rates, but not on marketshare predictions.

6.3. Quality of Bayesian Estimators

While calculating true Bayesian estimators takes too long to use in the question selection procedures, they can be computed offline from the questions and answers collected by each method. We evaluate the quality of two versions of such estimators. The first version is the true Bayesian estimator for each individual β^i for the model described in Section 3.2.2. While the prior distribution is the same Gaussian $\beta^i \sim N(\mu^0, \Sigma_0)$ for each individual (with a fixed population mean μ^0 and variance Σ_0 for each accuracy/heterogeneity regime), these estimators are computed independently using only the information collected for the specific individual. The second version is the estimator for a simple Hierarchical Bayesian (HB) model that considers all individuals at the same time. In this model, the individual β^i are also drawn from a common prior Gaussian distribution. However, the Gaussian prior is now $\beta^i \sim N(\nu^0, \Sigma_0)$ where the population mean ν^0 is now a random hyper-parameter whose prior distribution is $\nu^0 \sim N(\mu^0, \Sigma_0)$, where μ^0 and Σ_0 are the same parameters that are fixed for each accuracy/heterogeneity regime. Hence, this HB model uses the questions for all individuals to update the distribution of both the population mean ν^0 and all the individual partworths β^i. For both Bayesian models we sample from the posterior distributions of the corresponding parameters using STAN (Carpenter et al. 2016) through its Julia interface (Stan Development Team 2017). In all cases, we ran four MCMC chains for 5000 iterations plus the default 1000 warm-up iterations and evaluate their convergence using the standard Gelman-Rubin-Brooks potential scale reduction factor (PSRF) diagnostics provided by the Mamba.jl package (Smith 2017). Using the resulting samples (after dropping the warm-up ones) we compute the sample means and covariance matrices of the posterior distributions. We use the sample mean as the point-wise estimator and the sample covariance matrix to compute the D-Error of this estimator. Tables 3 and 4 present summaries of the quality measures for the individual Bayesian and Hierarchical Bayesian estimators respectively and in the same format as Table 2. In addition, Figure 4 presents the temporal evolution of these measures for the individual Bayesian estimator in the low accuracy and high heterogeneity regime (Figure 5 in Appendix B shows the same for the Hierarchical Bayesian estimators). Formatting for these figures is the same as for Figure 3.

We observe that the ellipsoidal method continues to provide a significant advantage on the quality of the point-wise estimator and hit rates for both the Bayesian and HB estimators. Also, as expected, the ellipsoid method provides a significant reduction of the D-Error of the point-wise estimator. Finally, we now have that the ellipsoid method also provides an advantage with regards to marketshare prediction errors. However, this advantage does not arises from a improvement of the marketshare predictions under the Bayesian and HB estimators for the ellipsoid method,

	Average D-Error			RMSE			Average Hit-Rate			Marketshare MAE		
# Questions	4	8	16	4	8	16	4	8	16	4	8	16
Low A, Low H												
Ellipsoidal	**0.22**	**0.19**	**0.15**	0.84	**0.76**	**0.67**	0.69	**0.70**	**0.72**	0.05	**0.05**	**0.04**
Polyhedral	0.24	0.23	0.21	0.87	0.84	0.80	0.68	0.69	0.70	0.05	0.05	0.05
Prob. Poly.	0.24	0.23	0.21	0.87	0.85	0.81	0.68	0.69	0.70	0.05	0.05	0.05
Robust	0.24	0.23	0.21	0.87	0.86	0.83	0.68	0.68	0.69	0.05	0.05	0.05
High A, Low H												
Ellipsoidal	**0.60**	**0.49**	**0.34**	0.50	**0.45**	**0.37**	0.81	**0.82**	**0.84**	**0.06**	**0.05**	**0.03**
Polyhedral	0.70	0.64	0.54	0.53	0.51	0.48	0.81	0.81	0.82	0.07	0.06	0.05
Prob. Poly.	0.70	0.64	0.54	0.53	0.51	0.47	0.80	0.81	0.82	0.07	0.06	0.05
Robust	0.69	0.64	0.54	0.53	0.51	0.46	0.80	0.81	0.82	0.07	0.06	0.05
Low A, High H												
Ellipsoidal	**0.79**	**0.62**	**0.41**	**1.10**	**0.98**	**0.76**	**0.68**	**0.71**	**0.76**	**0.07**	**0.05**	**0.04**
Polyhedral	0.90	0.78	0.61	1.19	1.12	0.97	0.65	0.67	0.71	0.08	0.06	0.05
Prob. Poly.	0.89	0.77	0.59	1.18	1.09	0.94	0.65	0.68	0.71	0.08	0.06	0.05
Robust	0.91	0.82	0.67	1.18	1.15	1.02	0.65	0.67	0.70	0.08	0.07	0.05
High A, High H												
Ellipsoidal	**2.24**	**1.67**	**0.96**	0.89	0.77	**0.55**	0.74	0.78	**0.84**	**0.10**	**0.07**	**0.04**
Polyhedral	2.46	2.01	1.36	0.89	0.82	0.67	0.74	0.76	0.81	0.11	0.08	0.05
Prob. Poly.	2.46	2.01	1.35	0.90	0.81	0.65	0.74	0.77	0.81	0.12	0.08	0.06
Robust	2.50	2.11	1.56	0.91	0.85	0.74	0.74	0.76	0.79	0.12	0.09	0.06

Table 3 Summary of Quality Measures for Individual Bayesian Estimator

	Average D-Error			RMSE			Average Hit-Rate			Marketshare MAE		
# Questions	4	8	16	4	8	16	4	8	16	4	8	16
Low A, Low H												
Ellipsoidal	**0.27**	**0.21**	**0.16**	0.87	**0.78**	**0.68**	0.68	**0.70**	**0.72**	0.07	**0.06**	0.05
Polyhedral	0.32	0.26	0.22	1.06	0.94	0.83	0.65	0.67	0.69	0.14	0.10	0.06
Prob. Poly.	0.32	0.26	0.22	0.97	0.88	0.82	0.66	0.68	0.69	0.12	0.06	0.05
Robust	0.35	0.28	0.23	0.88	0.92	0.87	0.68	0.67	0.68	**0.07**	0.08	0.07
High A, Low H												
Ellipsoidal	**0.71**	**0.53**	**0.34**	0.51	**0.46**	**0.37**	0.81	**0.82**	**0.84**	**0.07**	**0.05**	**0.03**
Polyhedral	0.82	0.69	0.56	0.56	0.53	0.48	0.79	0.80	0.82	0.09	0.07	0.05
Prob. Poly.	0.82	0.69	0.56	0.54	0.52	0.48	0.80	0.81	0.82	0.08	0.07	0.05
Robust	0.91	0.71	0.56	0.56	0.53	0.47	0.80	0.81	0.82	0.09	0.08	0.05
Low A, High H												
Ellipsoidal	**0.93**	**0.68**	**0.42**	1.19	**1.00**	**0.76**	0.66	**0.71**	**0.75**	0.13	**0.07**	**0.04**
Polyhedral	1.03	0.82	0.62	1.28	1.16	0.98	0.63	0.67	0.71	0.14	0.10	0.06
Prob. Poly.	1.02	0.81	0.60	1.28	1.13	0.94	0.63	0.67	0.72	0.14	0.11	0.05
Robust	1.27	0.90	0.69	1.19	1.19	1.02	0.64	0.66	0.70	**0.10**	0.09	0.07
High A, High H												
Ellipsoidal	**2.56**	**1.82**	**0.96**	0.90	0.78	**0.55**	0.74	0.78	**0.84**	**0.11**	0.08	**0.04**
Polyhedral	2.69	2.06	1.37	0.95	0.83	0.67	0.72	0.76	0.81	0.15	0.08	0.05
Prob. Poly.	2.70	2.07	1.37	0.93	0.82	0.65	0.73	0.76	0.81	0.14	0.08	0.06
Robust	3.18	2.25	1.59	0.93	0.87	0.75	0.73	0.76	0.79	0.13	0.08	0.06

Table 4 Summary of Quality Measures for Hierarchical Bayesian Estimator

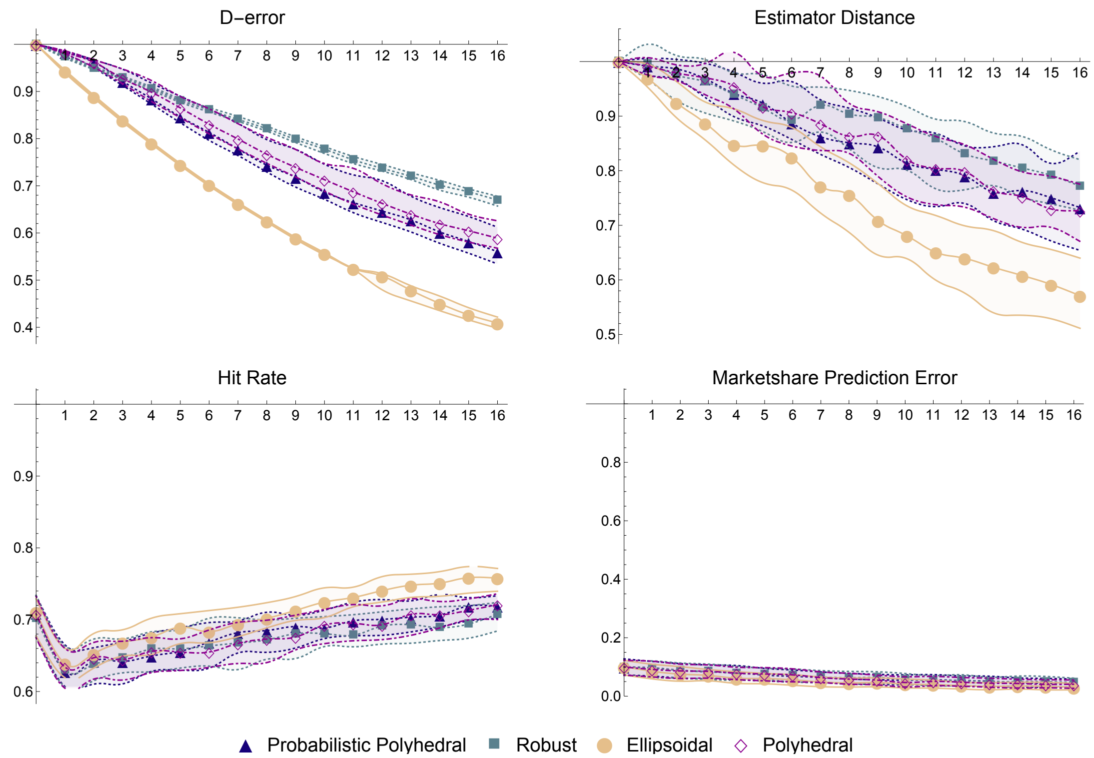


Figure 4 Temporal Evolution of Quality Measures for Individual Bayesian Estimator in the Low Accuracy and High Heterogeneity Regime

but from a deterioration of the predictions for the other methods. This is in contrast to a general improvement of the three other metrics for these methods under the Bayesian and HB estimators.

An additional interesting property of the ellipsoidal method is revealed when we compare all three estimators (the method's estimator, and the two Bayesian ones). It is apparent that there is little difference between the quality of ellipsoidal method's own estimator and the estimators recomputed using the individual or hierarchical Bayesian methods for the questions selected by the ellipsoidal method. Indeed, the RMSE (over all features, individuals, number of questions, accuracy and heterogeneity regimes) between the ellipsoidal method's own estimator and its individual Bayesian recalculation is 0.01 (cf. 0.49, 0.36 and 0.74 for the polyhedral, probabilistic polyhedral and robust methods respectively). Similarly, the RMSE for the hierarchical Bayesian recalculation is 0.16 (cf. 0.56, 0.45 and 0.77 for the polyhedral, probabilistic polyhedral and robust methods respectively). This suggests that the Bayesian approximations used by the ellipsoidal method in both the question selection and updates are quite accurate. Table 6 in Appendix B shows the temporal evolution of these RMSE through their evaluation right after the fourth, eighth and sixteenth questions.

6.4. Sensitivity to Prior Distribution

We next evaluate the relative sensitivity of the ellipsoidal method to the prior distribution not being centered at the true population mean. For this, we let the true population mean ν^0 be a random perturbation of the assumed population mean μ^0 such that $\|\mu^0 - \nu^0\| = \mathbb{E}_{\beta\sim N(\mu^0,\Sigma_0)} \|\mu^0 - \beta\|$. The true partworth vectors β^i are sampled from $N(\nu^0, \Sigma_0)$, but all methods and estimators still operate assuming they are samples from $N(\mu^0, \Sigma_0)$. Table 5 shows a summary for the usual metrics for this experiment in the low accuracy and high heterogeneity regime.

	Average D-Error			RMSE			Average Hit-Rate			Marketshare MAE		
# Questions	4	8	16	4	8	16	4	8	16	4	8	16
Method												
Ellipsoidal	**0.81**	**0.66**	**0.41**	1.02	**0.88**	**0.66**	**0.71**	**0.76**	**0.81**	**0.15**	**0.11**	**0.07**
Polyhedral	0.97	0.90	0.69	1.17	1.01	0.83	0.67	0.72	0.77	0.16	0.12	0.08
Prob. Poly.	0.90	0.82	0.64	1.07	0.96	0.80	0.69	0.73	0.77	0.18	0.12	0.08
Robust	0.95	0.90	0.79	1.35	1.33	1.25	0.62	0.62	0.66	0.20	0.18	0.15
Bayes												
Ellipsoidal	**0.79**	**0.62**	**0.40**	1.02	0.88	**0.66**	**0.71**	**0.76**	**0.81**	**0.15**	**0.10**	**0.07**
Polyhedral	0.90	0.78	0.61	1.06	0.95	0.80	0.69	0.73	0.77	0.18	0.14	0.10
Prob. Poly.	0.89	0.76	0.58	1.05	0.93	0.76	0.69	0.73	0.78	0.18	0.14	0.10
Robust	0.91	0.82	0.67	1.08	0.96	0.85	0.69	0.73	0.76	0.19	0.14	0.10
HB												
Ellipsoidal	1.09	**0.77**	**0.50**	0.94	0.82	**0.56**	0.75	0.79	**0.83**	0.13	0.09	**0.04**
Polyhedral	1.09	0.88	0.69	0.88	0.77	0.65	0.77	0.80	0.81	0.11	0.08	0.06
Prob. Poly.	1.09	0.87	0.66	**0.78**	0.73	0.65	0.78	0.80	0.81	**0.09**	0.08	0.06
Robust	1.31	0.97	0.75	1.08	0.91	0.71	0.71	0.78	0.80	0.18	0.11	0.08

Table 5 Sensitivity Analysis for Prior Distribution

We observe that the ellipsoidal method still outperforms the benchmark for most metrics.

7. Conclusions

In this paper we frame questionnaire design for choice-based conjoint analysis as a sequential decision-making problem under uncertainty. Considering a Bayesian approach, we have developed a near-optimal one-step look-ahead moment-matching approximate Bayesian policy for adaptive questionnaire design. We have formulated the one-step optimal question-selection as a MIP that can be solved effectively using state-of-the-art solvers.

We have shown that our method admits a simple and intuitive geometric interpretation, akin to that of the polyhedral method of Toubia et al. (2004). In this interpretation, the method maintains an ellipsoidal credibility region for the partworth vector, which is updated after each question. Questions are envisioned as hyperplanes on the partworth space, although unlike in the polyhedral method our update does not discard regions that are inconsistent with an answer, but instead adjusts the new credibility region to include the possibility of response-error.

The aforementioned geometric interpretation allows us to analyze optimal question-selection in terms of the traditional guidelines driving question-selection in previous methods. We show that optimal question-selection is achieved by striking a non-trivial trade-off between the choice-balance and the post-choice symmetry criteria, which can be computed and solved efficiently using optimization techniques. Our numerical experiments have shown how incorporating such insight into existing methods achieves a significant improvement in various performance metrics, including the D-error, RMSE and out of sample hit-rates.

Our formulation of the questionnaire design problem assumes that consumer-choice follows a multinomial Logit model. The results and methods can be extended directly to different specifications of the distribution of the idiosyncratic shock to utility, as long as said shocks remain uncorrelated among profiles. Indeed, under a different noise specification, one would only have to adjust the distribution of the Z random variables in Propositions 1 and 2, on top of minor adjustments to computations throughout our algorithms. However, the efficiency of our multivariate normal approximation to the posterior distribution, and therefore of the whole method, would be affected by the new specification. Consider for example the trivial case of no response-error: the exact posterior computation would be that of the polyhedral method, and the multivariate normal approximation may not be reasonable. In this regard, the suitability of our approach would depend on the relative variance of the idiosyncratic shock relative to that of the prior distribution.

Our results so far apply to the case of two profiles per question, which might include a *no purchase* option (for this, one simply imposes some structure on the set $\mathcal{X}$). Most of the proposed framework adapts directly to the case of more profiles per question; one needs to be especially careful when specifying the approximate posterior update and the question-selection MIP. In the more general case of r profiles, the answer to a question can be seen as inserting $r-1$ hyperplanes through a credibility region in the geometric interpretation of the method. However, the fast ellipsoidal update in Proposition 1 would have to be adjusted: from the proof of Proposition 1, we see that one would require the computation of the mean and covariance matrix of a $r-1$ dimensional random vector (Z). Computationally, this will replace the one-dimensional integration for the case of two profiles with a $r-1$ dimensional integration computation, which becomes more expensive as r grows larger. Similarly, optimal question-selection would have to be modified along the same lines. That is, a new function g would arise, which would depend on the distances of the many hyperplanes to the center of the prior mean, and on the covariance matrix associated with said hyperplanes (interpreted as in Section 5.5). This would increase the computational cost of the procedure, but remains practical for a small but reasonable number of profiles. Note that, in our current implementation, question-selection as well as update computations can be performed with no noticeable delay, thus allowing the implementation of the method in real-time.

With regards to the scalability of the method we should consider the dimension of the partworth vector and the size of the grid used to construct piecewise linear function $\breve{g}$. With regards to the later, we did not detect significant sensitivity to the grid size or resolution in our experiments (reducing it did not significantly reduce the performance of the method and increasing it did not significantly increase the solve times for question-selection). However, it is possible to construct instances where a high grid resolution is needed to maintain the ellipsoidal method's advantage. Fortunately, the size of advanced MIP formulations for piecewise linear functions grow very slowly (logarithmically) with this resolution and can hence easily handle large grids (Huchette and Vielma 2017, Vielma et al. 2010). In contrast, while the numerical integration time needed to evaluate D-error function $g(m,v)$ at a single point (m,v) is extremely small, the number of evaluations needed to build $\breve{g}$ for a $k \times k$ grid is k^2, which could lead to an unacceptable question-selection time for high resolutions. In such cases, a more tractable approach would be to build a piecewise linear approximation of the Fisher-based approximation of D-error described in Proposition 3, which does not require numerical integration. To evaluate the effectiveness of such an approach we repeated our computational experiments with exactly such a variant of the ellipsoid method and obtained an almost identical behavior to the original ellipsoidal method. Detailed results of this experiment are included in Appendix B.3. With regards to scalability on the dimension n of the partworth vector, we note that the size of (9) grows as n^2 because of the standard approach to linearize products of binary variables. Hence, working with very large n will likely require applying more advanced techniques in this part of the formulation.

Finally, the application of the Bayesian ellipsoidal method requires a working prior distribution over the partworth vector, which might lack in practice. In this regard, practical application of the method might required a two stage procedure, akin the steps in existing methods, e.g. the aggregate customization approach of Arora and Huber (2001). Ongoing research aims at that direction, and on the application of the method in field experiments. The results in this paper can be seen as providing both the theoretical and technological grounds to test the method in live conjoint experiments, which is the subject of current research.

Endnotes

1. Within a random utility framework, a response-error is said to occur whenever a respondent does not select the alternative that would be chosen absent idiosyncratic shocks to utility; see Section 3.2.1

2. Indeed, if we consider a uniform distribution over the ellipsoidal region, its mean is equal to μ^0 and its covariance matrix is equal to Σ_0, which are precisely the mean and covariance matrix of the multivariate normal prior.

3. $\mathbb{E}_a\left(\mathrm{tr}\left(Cov\left(\beta|a\right)\right)\right)$, $\mathbb{E}_a\left(c^\top Cov\left(\beta|a\right)c\right)$, and $\mathbb{E}_a\left(\lambda_{\max}\left(Cov\left(\beta|a\right)\right)\right)$ respectively, where $\mathrm{tr}(\cdot)$ is the trace of a matrix, c is an arbitrary fixed vector and $\lambda_{\max}(\cdot)$ is the largest eigenvalue of a matrix.

4. Cutting off-center yields one half-ellipsoid with a smaller volume and one with a larger volume, but the resultant low probability of the respondent picking the first one offsets any potential gains here.

5. The prior ellipsoid for Figure 2 is taken to be the 90% credibility ellipsoid for a zero-mean multivariate normal with diagonal covariance matrix and $(\Sigma_0)_{1,1} = 100$ and $(\Sigma_0)_{2,2} = 1$ (the aspect ratio of the figure is modified to improve the aesthetics).

6. Indeed, Huber et al. (1993) find such an inconsistency when ranking methods based on these two criteria (Hit-rate and marketshare predictions) and propose an approach to calculate marketshare (which we use in our tables) that mitigates, but does not necessarily eliminate this inconsistency. For more details and a partial explanation for this phenomenon we refer the reader to Elrod and Kumar (1993).

Acknowledgments

The authors are grateful to the Department Editor, Associate Editor, and the referees for constructive comments that helped improve the paper. In particular, we thank the Associate Editor for pointing out that a Fisher-based version of D-error can also be written as a function of only the mean and variance of the target question. The second author's research was partially funded by project FONDECYT 11140261 and the Complex Engineering Systems Institute, ISCI (ICM-FIC: P05-004-F, CONICYT: FB0816)

References

Abernethy, J., Evgeniou, T., Toubia, O. and Vert, J. (2008), 'Eliciting consumer preferences using robust adaptive choice questionnaires', *IEEE Transactions on Knowledge and Data Engineering* **20**(2), 145–155.

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, second edn, Wiley, New York, NY.

Arora, N. and Huber, J. (2001), 'Improving parameter estimates and model prediction by aggregate customization in choice experiments', *Journal of Consumer Research* **28**(2), 273–283.

Bellman, R. (1957), *Dynamic Programming*, 1 edn, Princeton University Press, Princeton, NJ, USA.

Bertsimas, D. and O'Hair, A. (2013), 'Learning preferences under noise and loss aversion: An optimization approach', *Operations Research* **61**(5), 1190–1199.

Bertsimas, D. and Tsitsiklis, J. (1997), *Introduction to Linear Optimization*, 1st edn, Athena Scientific.

Bezanson, J., Edelman, A., Karpinski, S. and Shah, V. B. (2017), 'Julia: A fresh approach to numerical computing', *SIAM Review* **59**(1), 65–98.

Bliemer, M. C. and Rose, J. M. (2010), 'Construction of experimental designs for mixed logit models allowing for correlation across choice observations', *Transportation Research Part B: Methodological* **44**(6), 720 – 734.

Boyd, S. and Vandenberghe, L. (2004), *Convex optimization*, Cambridge university press.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P. and Riddell, A. (2016), 'Stan: A probabilistic programming language', *Journal of Statistical Software* **20**, 1–37.

Cohen, M., Lobel, I. and Paes Leme, R. (2016), Feature-based dynamic pricing. Working paper, New York University.

Crabbe, M., Akinc, D. and Vandebroek, M. (2014), 'Fast algorithms to generate individualized designs for the mixed logit choice model', *Transportation Research Part B: Methodological* **60**, 1–15.

Cui, D. and Curry, D. (2005), 'Prediction in marketing using the support vector machine', *Marketing Science* **24**(4), 595–615.

Dunning, I., Huchette, J. and Lubin, M. (2017), 'JuMP: a modeling language for mathematical optimization', *SIAM Review* **59**(2), 295–320.

Dzyabura, D. and Hauser, J. R. (2011), 'Active machine learning for consideration heuristics', *Marketing Science* **30**(5), 801–819.

Elrod, T. and Kumar, S. K. (1993), Bias in the first choice rule for predicting share, *in* 'Proceedings of the 1998 Sawtooth Software Conference', pp. 259–271.

Evgeniou, T., Boussios, C. and Zacharia, G. (2005), 'Generalized robust conjoint estimation', *Marketing Science* **24**(3), 415–429.

Evgeniou, T., Pontil, M. and Toubia, O. (2007), 'A convex optimization approach to modeling consumer heterogeneity in conjoint estimation', *Marketing Science* **26**(6), 805–818.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. and Rubin, D. (2013), *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.

Green, P. E. and Srinivasan, V. (1978), 'Conjoint analysis in consumer research: Issues and outlook', *Journal of Consumer Research* **5**(2), 103–23.

Grötschel, M., Lovasz, L. and Schrijver, A. (2011), *Geometric Algorithms and Combinatorial Optimization*, Algorithms and Combinatorics, Springer Berlin Heidelberg.

Harville, D. A. (2000), *Matrix Algebra From a Statistician's Perspective*, Springer.

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

Huang, D. and Luo, L. (2016), 'Consumer preference elicitation of complex products using fuzzy support vector machine active learning', *Marketing Science* **35**(3), 445–464.

Huber, J., Fiedler, J. and Miller, R. (1993), 'The effectiveness of alternative preference procedures in predicting choice', *Journal of Marketing Research* **30**(1), 105–114.

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

Huchette, J. (2017), 'PiecewiseLinearOpt.jl', `https://github.com/joehuchette/PiecewiseLinearOpt.jl`. Accessed: 2017-08-15.

Huchette, J. and Vielma, J. P. (2017), 'Nonconvex piecewise linear functions: Advanced formulations and simple modeling tools', *arXiv:1708.00050* .

IBM ILOG (2015), 'CPLEX High-performance mathematical programming engine'. `http://www.ibm.com/software/integration/optimization/cplex/`.

Kanninen, B. J. (2002), 'Optimal design for multinomial choice experiments', *Journal of Marketing Research* **39**(2), 214–227.

Kessels, R., Jones, B., Goos, P. and Vandebroek, M. (2009), 'An efficient algorithm for constructing bayesian optimal choice designs', *Journal of Business & Economic Statistics* **27**(2), 279–291.

Kuhfeld, W., Tobias, R. and Garrat, M. (1994), 'Efficient experimental design with marketing research applications', *Journal of Marketing Research* **31**(4), 545–575.

Liu, Q. and Tang, Y. (2015), 'Construction of heterogeneous conjoint choice designs: A new approach', *Marketing Science* **34**(3), 346–366.

Louviere, J. J., Hensher, D. A. and Swait, J. D. (2000), *Stated Choice Methods: Analysis and Application*, Cambridge University Press, New York, NY, USA.

Lubin, M. and Dunning, I. (2015), 'Computing in operations research using julia', *INFORMS Journal on Computing* **27**(2), 238–248.

McFadden, D. (1974), Conditional logit analysis of qualitative choice behaviour, *in* P. Zarembka, ed., 'Frontiers in Econometrics', Academic Press New York, New York, NY, USA, pp. 105–142.

Powell, W. B. (2007), *Approximate dynamic programming : solving the curses of dimensionality*, Wiley series in probability and statistics, Hoboken, N.J. Wiley-Interscience.

QuadGK.jl (2017), `https://github.com/JuliaMath/QuadGK.jl`. Accessed: 2017-08-15.

Sándor, Z. and Wedel, M. (2001), 'Designing conjoint choice experiments using managers' prior beliefs', *Journal of Marketing Research* **38**(4), 430–444.

Sándor, Z. and Wedel, M. (2005), 'Heterogeneous conjoint choice designs', *Journal of Marketing Research* **42**(2), 210–218.

Sauré, D. and Vielma, J. P. (2017), 'Source code for "Ellipsoidal methods for adaptive choice-based conjoint analysis"', `https://github.com/juan-pablo-vielma/EMFACBCA`. Accessed: 2017-11-19.

Segall, D. O. (2010), Principles of multidimensional adaptive testing, *in* W. J. van der Linden and C. A. W. Glas, eds, 'Elements of Adaptive Testing', Springer, New York, NY.

Silvey, S. (1980), *Optimal Design: An Introduction to the Theory for Parameter Estimation*, Springer, New York, NY.

Smith, B. J. (2017), 'Mamba: Markov chain monte carlo (MCMC) for bayesian analysis in julia', `https://github.com/brian-j-smith/Mamba.jl`. Accessed: 2017-08-15.

Sonnevend, G. (1986), An "analytical centre" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming, *in* A. Prékopa, J. Szelezsáan and B. Strazicky, eds, 'System Modelling and Optimization: Proceedings of 12th IFIP Conference, Budapest, Hungary, September 2–6, 1985', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 866–875.

Stan Development Team (2017), 'Stan.jl: the julia interface to stan', `https://github.com/goedman/Stan.jl`. Accessed: 2017-08-15.

Toubia, O., Hauser, J. and Garcia, R. (2007), 'Probabilistic polyhedral methods for adaptive choice-based conjoint analysis: Theory and application', *Marketing Science* **26**(5), 596–610.

Toubia, O., Hauser, J. R. and Simester, D. I. (2004), 'Polyhedral methods for adaptive choice-based conjoint analysis', *Journal of Marketing Research* **41**(1), 116–131.

Toubia, O., Johnson, E., Evgeniou, T. and Delquié, P. (2013), 'Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters', *Management Science* **59**(3), 613–640.

Vielma, J. P., Ahmed, S. and Nemhauser, G. (2010), 'Mixed-integer models for nonseparable piecewise linear optimization: unifying framework and extensions', *Operations Research* **58**, 303–315.

Yu, J., Goos, P. and Vandebroek, M. (2011), 'Individually adapted sequential bayesian conjoint-choice designs in the presence of consumer heterogeneity', *International Journal of Research in Marketing* **28**(4), 378–388.

Yu, J., Goos, P. and Vandebroek, M. (2012), 'A comparison of different bayesian design criteria for setting up stated preference studies', *Transportation Research Part B: Methodological* **46**(7), 789–807.

Yu, J., Goss, P. and Vanderbroek, M. (2009), 'Efficient conjoint choice designs in the presence of respondent heterogeneity', *Marketing Science* **28**(1), 122–135.

Appendix A: Omitted Proofs

Proof of Proposition 1. Define $v := x - y$ and note that $\beta \sim \mu + \Sigma^{1/2} Z$, where $Z \sim N(0, I)$ and $I \in \mathbb{R}^{n\times n}$ denotes the identity matrix. Consider an orthonormal basis $\{w_1, \ldots, w_n\}$ of $\mathbb{R}^n$ such that

$$w_1 := \left(\Sigma^{1/2}\right)^\top v/\sigma_{x,y}.$$

Thus, we have that $\beta \sim \mu + \Sigma^{1/2} W Z$ as well. Let us focus on the posterior distribution $f_{Z|\Omega}(\cdot)$ of Z given Ω. Per Bayes' rule, we have that

$$\begin{aligned}
f_{Z|\Omega}(z) &= \frac{\mathbb{P}(\Omega|z)}{\mathbb{P}(\Omega)} \prod_{i=1}^{n} \phi(z_i) \\
&= \frac{\mathbb{P}\left(\mu_{x,y} + \Sigma^{1/2} W z \cdot v + \varepsilon_x \geq \varepsilon_y\right) \prod_{i=1}^{n} \phi(z_i)}{\int_{\mathbb{R}^n} \mathbb{P}\left(\mu_{x,y} + \Sigma^{1/2} W z' \cdot v + \varepsilon_x \geq \varepsilon_y\right) \prod_{i=1}^{n} \phi(z'_i) dz'_i} \\
&= \frac{\left(1 + \exp\left(-\mu_{x,y} - v \cdot \Sigma^{1/2} W z\right)\right)^{-1} \prod_{i=1}^{n} \phi(z_i)}{\int_{\mathbb{R}^n} \left(1 + \exp\left(-\mu_{x,y} - v \cdot \Sigma^{1/2} W z'\right)\right)^{-1} \prod_{i=1}^{n} \phi(z'_i) dz'_i} \\
&= \frac{\left(1 + \exp\left(-\mu_{x,y} - \sigma_{x,y} z_1\right)\right)^{-1} \prod_{i=1}^{n} \phi(z_i)}{\int_{\mathbb{R}^n} \left(1 + \exp\left(-\mu_{x,y} - \sigma_{x,y} z'_1\right)\right)^{-1} \prod_{i=1}^{n} \phi(z'_i) dz'_i} \\
&= C \frac{\phi(z_1)}{1 + \exp\left(-\mu_{x,y} - \sigma_{x,y} z_1\right)} \prod_{i=2}^{n} \phi(z_i),
\end{aligned}$$

where $C := \left(\int_{\mathbb{R}} \left(1 + \exp\left(-\mu_{x,y} - \sigma_{x,y} z'\right)\right)^{-1} \phi(z') dz' \right)^{-1}$. We conclude that the components of the posterior z are independent, and that $Z_i \sim N(0,1)$ for all $i > 1$. Consider now a random variable Z' with density

$$f_{Z'}(z') := C \left(1 + e^{-\mu_{x,y} - \sigma_{x,y} z'}\right)^{-1} \phi(z').$$

Because $\beta \sim \mu + \Sigma^{1/2} W Z$, we have that

$$\mathbb{E}\{\beta|\Omega\} = \mu + \left(\Sigma^{1/2} W\right)_1 \mathbb{E}(Z'), \quad \text{and} \quad Cov(\beta|\Omega) = \Sigma^{1/2} W \begin{pmatrix} \mathrm{Var}(Z') & 0 \\ 0 & I' \end{pmatrix} \left(\Sigma^{1/2} W\right)^{\top},$$

where $I' \in \mathbb{R}^{n-1 \times n-1}$ denotes the identity matrix. This concludes the proof.

Proof of Proposition 2. Set $m = \mu_{x,y}$ and $v = \sigma_{x,y}$. Using Proposition 1, we have that

$$Cov(\beta|a=1) = \left(\Sigma^{1/2} W\right) \begin{pmatrix} Var(Z(m,v)) & 0 \\ 0 & I' \end{pmatrix} \left(\Sigma^{1/2} W\right)^{\top}.$$

Exchanging the roles of x and y in Proposition 1, we have that

$$Cov(\beta|a=2) = \left(\Sigma^{1/2} W\right) \begin{pmatrix} Var(Z(-m,v)) & 0 \\ 0 & I' \end{pmatrix} \left(\Sigma^{1/2} W\right)^{\top}.$$

From the fact that W is an orthonormal basis, we have that

$$\det\left(Cov(\beta|a=1)\right) = \det(\Sigma)\ Var(Z(m,v)) \quad \text{and} \quad \det\left(Cov(\beta|a=2)\right) = \det(\Sigma)\ Var(Z(-m,v)).$$

From the above, and noting that $\mathbb{P}(a=1)=p(m,v)=1-p(-m,v)$, we obtain that

$$\begin{aligned}\mathbb{E}_a\left(\det\left(Cov(\beta|a)\right)^{1/d}\right) &= p(m,v)\ \det\left(Cov(\beta|a=1)\right)^{1/d}+p(-m,v)\ \det\left(Cov(\beta|a=2)\right)^{1/d}\\ &= \det\left(\Sigma\right)^{1/d}\left(p(m,v)Var(Z(m,v))^{1/d}+p(-m,v)Var(Z(-m,v))^{1/d}\right).\end{aligned}$$

The result follows from noting that the above depends on (x,y) only through $m=\mu_{x,y}$ and $v=\sigma_{x,y}$.

Proof of Proposition 3. The first result follows directly from the proof of Proposition 2. For the second result, from (10) we have that

$$\det\left(H(\alpha)\right)=\det\left(\left(1+e^{-\alpha(x-y)}\right)^{-1}\left(1+e^{\alpha(x-y)}\right)^{-1}(x-y)\,(x-y)^{\top}+\Sigma^{-1}\right)^{-1}.$$

Then, by the matrix-determinant lemma (Harville 2000), we have that

$$\det\left(H(\alpha)\right)=\det\left(\Sigma\right)\left(1+\frac{\sigma_{x,y}^2}{2+e^{-\alpha(x-y)}+e^{\alpha(x-y)}}\right)=\det\left(\Sigma\right)\left(1+\frac{\sigma_{x,y}^2}{2+2\ \cosh(\alpha(x-y))}\right).$$

The result follows from evaluating at $\alpha=\mu$ and noting that $\mu\cdot(x-y)=\mu_{x,y}$.

For the third result, note that if $\alpha\cdot(x-y)\sim\mu_{x,y}+\sigma_{x,y}\ Z$, with $Z\sim N(0,1)$, then

$$\mathbb{E}_{\alpha\sim N(\mu,\Sigma)}\left(\psi\left(\det\left(H(\alpha)\right)^{-1}\right)\right)=\int_{\mathbb{R}}\psi\left(\det\left(\Sigma\right)^{-1}\left(1+\frac{\sigma_{x,y}^2}{2+2\cosh\left(\mu_{x,y}+\sigma_{x,y}z\right)}\right)^{-1}\right)\phi(z)dz.$$

For the last result note that from Section 3.1 in Grötschel et al. (2011) we have that if $n(2a-3)\mu_{x,y}/\sigma_{x,y}<1$ then $\mathcal{M}(a)=\mathcal{E}\left(\overline{\mu},\overline{\Sigma},r\right)$ for

$$\overline{\mu}=\mu-\frac{(2a-3)\left(1-n(2a-3)\mu_{x,y}/\sigma_{x,y}\right)\Sigma(y-x)}{(n+1)\sigma_{x,y}}$$

and

$$\overline{\Sigma}=\left(\frac{n^2\left(1-\mu_{x,y}^2/\sigma_{x,y}^2\right)}{n^2-1}\right)\left(\Sigma-\frac{2\left(1-n(2a-3)\mu_{x,y}/\sigma_{x,y}\right)}{\sigma_{x,y}^2(n+1)\left(1-(2a-3)\mu_{x,y}/\sigma_{x,y}\right)}\Sigma(x-y)(x-y)^{\top}\Sigma\right).$$

If instead we have $n(2a-3)\mu_{x,y}/\sigma_{x,y}\geq 1$, then $\mathcal{M}(a)=\mathcal{E}\left(\mu,\Sigma,r\right)$. Then,

$$\mathrm{vol}\left(\mathcal{M}(a)\right)=\begin{cases}\frac{\left(\det\left(\overline{\Sigma}\right)\right)^{1/2}r^n\pi^{n/2}}{\Gamma(n/2+1)} & n(2a-3)\mu_{x,y}/\sigma_{x,y}<1\\ \frac{\left(\det(\Sigma)\right)^{1/2}r^n\pi^{n/2}}{\Gamma(n/2+1)} & n(2a-3)\mu_{x,y}/\sigma_{x,y}\geq 1\end{cases}$$

and by the matrix determinant lemma we have that $\det\left(\overline{\Sigma}\right) = \det\left(\Sigma\right)\overline{g}(\mu_{x,y},\sigma_{x,y},n,a)$ for

$$\overline{g}(m,v,n,a) := \left(\frac{n^2\left(1-m^2/v^2\right)}{n^2-1}\right)^n \left(1-\frac{2\left(1-n(2a-3)m/v\right)}{v(n+1)\left(1-(2a-3)m/v\right)}\right).$$

Then the result follows for

$$g\left(m,v,n\right) := \frac{r^n\pi^{n/2}}{\Gamma\left(n/2+1\right)}\left(p\left(m,v\right)\left(\hat{g}(m,v,n,1)\right)^{1/2} + p\left(-m,v\right)\left(\hat{g}(m,v,n,2)\right)^{1/2}\right),$$

where $p\left(m,v\right)$ is as defined in Proposition 2 and

$$\hat{g}(m,v,n,a) := \begin{cases} \overline{g}(m,v,n,a) & n(2a-3)\mu_{x,y}/\sigma_{x,y} < 1 \\ 1 & n(2a-3)\mu_{x,y}/\sigma_{x,y} \geq 1 \end{cases}.$$

Appendix B: Additional Graphs and Tables

B.1. Hierarchical Bayesian Estimator

Figure 5 presents a variant of Figure 4 that uses the Hierarchical Bayesian estimator instead of the individual one.

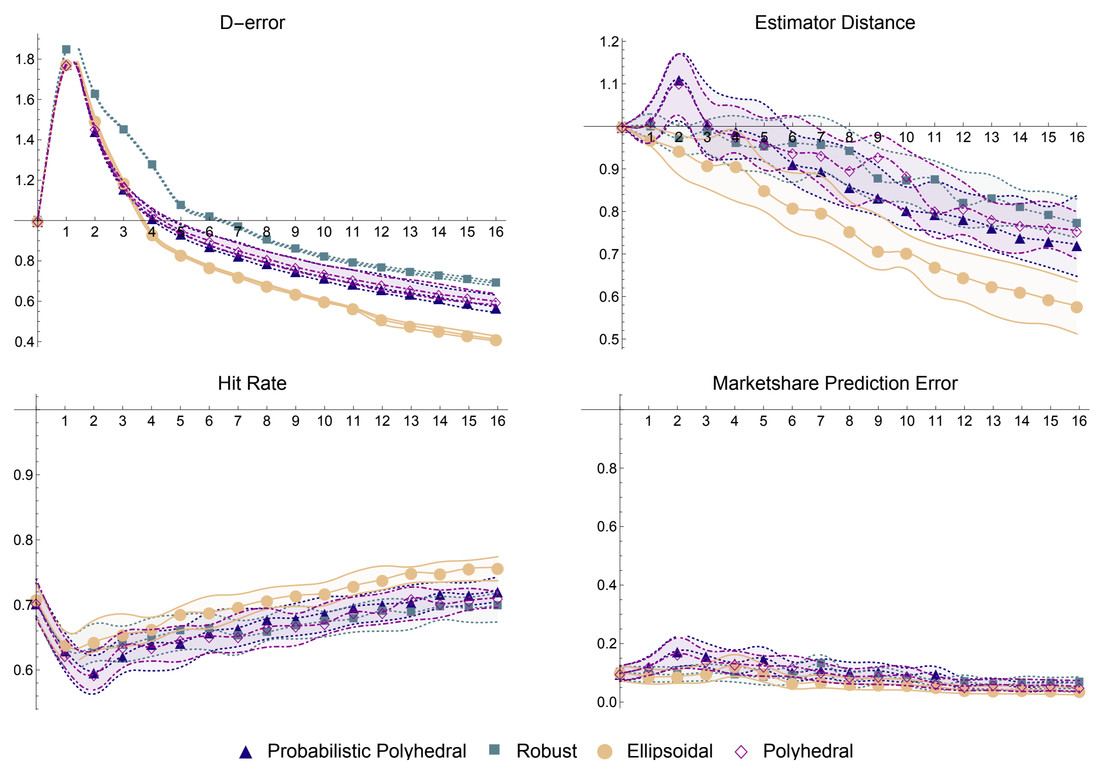


Figure 5 Temporal Evolution of Quality Measures for Hierarchical Bayesian Estimator in the Low Accuracy and High Heterogeneity Regime

B.2. Quality of the Each Methods Estimator

Table 6 presents the RMSE between each methods estimator and the corresponding individual and hierarchical Bayesian recalculations.

	M-B RMSE			M-HB RMSE		
# Questions	4	8	16	4	8	16
High A, Low H						
Ellipsoidal	0.01	0.01	0.01	0.26	0.18	0.08
Polyhedral	0.67	0.65	0.66	0.86	0.76	0.68
Prob. Poly.	0.30	0.48	0.64	0.51	0.51	0.65
Robust	0.63	0.79	0.75	0.68	0.83	0.79
Low A, High H						
Ellipsoidal	0.00	0.00	0.01	0.12	0.07	0.05
Polyhedral	0.38	0.36	0.36	0.41	0.37	0.36
Prob. Poly.	0.24	0.25	0.31	0.25	0.28	0.32
Robust	0.39	0.48	0.46	0.42	0.50	0.46
High A, High H						
Ellipsoidal	0.01	0.01	0.03	0.53	0.20	0.09
Polyhedral	0.61	0.48	0.40	0.81	0.55	0.39
Prob. Poly.	0.23	0.32	0.32	0.59	0.49	0.32
Robust	0.91	1.11	1.04	0.95	1.17	1.05
Low A, Low H						
Ellipsoidal	0.01	0.01	0.04	0.17	0.16	0.07
Polyhedral	0.45	0.32	0.17	0.55	0.31	0.17
Prob. Poly.	0.27	0.19	0.13	0.38	0.18	0.14
Robust	0.70	0.80	0.69	0.72	0.84	0.71

Table 6 RMSE between each method's estimator and the individual Bayesian and Hierarchical Bayesian estimators

B.3. Question Selection

While the one-dimensional numerical integration needed to compute D-error through g in Proposition 2 is extremely fast, there may be settings in which avoiding this may be desirable. For this reason we now compare the base ellipsoidal method with a variant that use the Fisher-based approximations of D-error (11) from Proposition 3 with $\psi(\cdot) = (\cdot)^{1/n}$. We denote this variant simply as *Fisher*. As a reference, we also include a variant that uses the alternative Fisher-based approximations of D-error (12) from Proposition 3 also with $\psi(\cdot) = (\cdot)^{1/n}$. This alternative approximation does require one-dimensional numerical integration and we denote it as *Expected Fisher*. Finally, we also include a variant that uses the heuristic question selection method from the polyhedral method (i.e. the heuristics aim of maximizing the question's variance and minimizing the absolute value of it's mean). We denote this variant as *MaxMin*. Table 7 shows the usual performance metrics for the individual Bayesian estimator derived from the questions generated by all four variants (including

	Average D-Error			RMSE			Average Hit-Rate			Marketshare MAE		
# Questions	4	8	16	4	8	16	4	8	16	4	8	16
Method												
Fisher	0.81	0.66	0.42	1.12	1.00	0.75	0.67	0.71	0.76	0.07	0.05	0.04
Expected Fisher	0.81	0.66	0.42	1.12	1.00	0.77	0.67	0.71	0.76	0.07	0.05	0.04
Ellipsoidal	0.81	0.66	0.42	1.10	0.98	0.76	0.68	0.71	0.76	0.07	0.05	0.04
MaxMin	0.91	0.82	0.66	1.16	1.11	1.01	0.65	0.67	0.70	0.09	0.07	0.05
Bayes												
Fisher	0.79	0.62	0.41	1.12	1.00	0.75	0.67	0.71	0.76	0.07	0.05	0.04
Expected Fisher	0.79	0.62	0.41	1.12	1.00	0.77	0.67	0.70	0.76	0.07	0.06	0.04
Ellipsoidal	0.79	0.62	0.41	1.10	0.98	0.76	0.68	0.71	0.76	0.07	0.05	0.04
MaxMin	0.92	0.84	0.68	1.16	1.11	1.01	0.65	0.67	0.70	0.09	0.07	0.05
HB												
Fisher	0.93	0.68	0.41	1.16	1.00	0.75	0.66	0.71	0.76	0.10	0.06	0.03
Expected Fisher	0.92	0.68	0.42	1.17	1.01	0.78	0.66	0.70	0.76	0.11	0.07	0.04
Ellipsoidal	0.93	0.68	0.42	1.19	1.00	0.76	0.66	0.71	0.75	0.13	0.07	0.04
MaxMin	1.18	0.92	0.70	1.26	1.18	1.01	0.63	0.65	0.70	0.18	0.16	0.07

Table 7 Impact of the question selection procedure on the ellipsoidal method.

the original ellipsoidal method). We can see that both Fisher-based approximations perform almost identically to the original ellipsoidal method. Hence, the first one (Fisher) is a viable alternative when avoiding one-dimensional numerical integration is prefered. Furthermore, as expected, all three approximate-optimal question selection procedures provide an advantage over the MaxMin heuristic. Finally, we note that all four methods use the same moment-matching update of the ellipsoidal method and their only difference is how they select questions. Hence, all four methods yield estimators that are extremely close to the individual Bayesian estimator (cf. Table 6) and the alternative version of Table 7 using method's estimator is extremely similar to the original table.