Properties of Preference Questions: Utility Balance, Choice Balance, Configurators, and M-Efficiency

by

John R. Hauser

and

Olivier Toubia

August 21, 2003

John R. Hauser is the Kirin Professor of Marketing, Sloan School of Management, Massachusetts Institute of Technology, E56-314, 38 Memorial Drive, Cambridge, MA 02142, (617) 253-2929, fax (617) 253-7597, jhauser@mit.edu.

Olivier Toubia is a graduate student at the Marketing Group, Massachusetts Institute of Technology, E56-345, 38 Memorial Drive, Cambridge, MA 02142, (617) 253-7353, fax (617) 253-7597, toubia@mit.edu.

This research was supported by the MIT Sloan School of Management and MIT's Center for Innovation in Product Development. This paper may be downloaded from the Virtual Customer Initiative's website, mitsloan.mit.edu/vc. This website also contains demonstrations of many of the preference measurement questions discussed in this paper, including configurators. Special thanks to Eric Bradlow, Theodoros Evgeniou, and Duncan Simester for insightful comments and Ely Dahan who was instrumental in developing the configurator with which we collected the data for the stylized fact on warm-up questions. Limor Weisberg provided the web design for the configurator and the examples of paired-comparison and choice-based questions. The authors are listed in alphabetical order. Contributions were equal and synergistic.

Properties of Preference Questions: Utility Balance, Choice Balance, Configurators, and M-Efficiency

Abstract

This paper uses stylized models to explore efficient preference questions. The formal analyses provide insights on criteria that are common in widely-used question-design algorithms. We demonstrate that the criterion of metric utility balance leads to inefficient questions, absolute biases, and relative biases among partworths. On the other hand, choice balance for stated-choice questions (the analogy of utility balance for metric questions) can lead to efficiency under some conditions, especially for discrete features. When at least one focal feature is continuous, choices are imbalanced at optimal efficiency. This choice imbalance declines as responses become more accurate and non-focal features increase. The analysis of utility- and choice-balance provide insights on the relative success of metric- and choice-based adaptive polyhedral methods and suggest future improvements.

We then use the formal models to explore a new form of data collection that has become popular in e-commerce and web-based market research – configurators. We study the stylized fact that questions normally used as warm-up questions in conjoint analysis are surprising accurate in predicting consumer behavior that is relevant to e-commerce applications. For these applications, new data suggest that warm-up questions are, perhaps, more useful managerially than paired-comparison tradeoff questions. We examine and generalize this insight to demonstrate the advantages of matching tradeoff, configurator, and "range" questions to the relevant managerial decisions. This further generalizes to M-efficiency (managerial efficiency), an alternative criteria for question design algorithms. Designs which minimize D-errors and A-errors may not minimize M_D- and M_A-errors which measure precision relevant to managerial decisions. We provide examples in which simple modifications to question design increase M-efficiency.

Keywords: Conjoint analysis, efficient question design, adaptive question design, Internet market research, e-commerce, product development.

1. Motivation

Preference measurement is crucial to managerial decisions in marketing. Methods such as conjoint analysis, voice-of-the-customer methods, and, more recently, configurators, are used widely for product development, advertising, and strategic marketing. In response to this managerial need, marketing scientists have developed many sophisticated and successful algorithms to select questions efficiently. For example, Figure 1 illustrates two popular formats that have been studied widely.

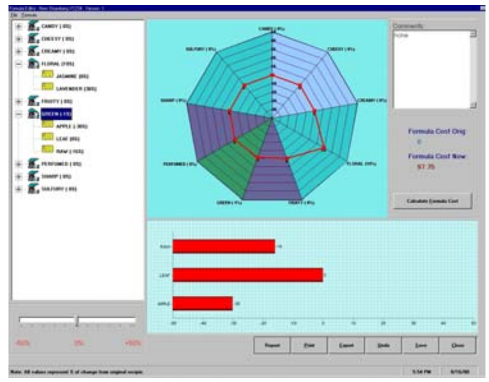
Figure 1
Popular Formats for Conjoint Analysis

(a) Metric paired-comparison format (laptop bags) (b) Stated-choice format (Executive Education)

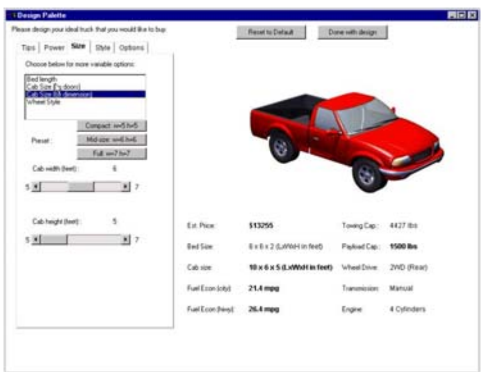
For the metric-pairs format (Figure 1a) questions can be selected with efficient designs, with Adaptive Conjoint Analysis (ACA), or with polyhedral methods (Kuhfield, Tobias, and Garratt 1994; Sawtooth 2002; Toubia, Simester, Hauser, Dahan 2003). For the stated-choice format (Figure 1b), recent advances adapt questions based on pretests, managerial beliefs, or prior questions and do so either by selecting an aggregate design that is replicated for all subsequent respondents or by adapting question design for each respondent (Arora and Huber 2001; Huber and Zwerina 1996; Johnson, Huber and Bacon 2003; Kanninen 2002; Sandor and Wedel 2001, 2002, Toubia, Hauser, and Simester 2003). These question-design methods are used with a variety of estimation methods (regression, Hierarchical Bayes estimation, mixed logit models, analytic center approximations, and hybrid methods) and have proven to provide more efficient and more accurate data with which to estimate partworths.

Figure 2
Example Configurators

(a) Configurator from Dell.com



(b) Innovation Toolkit for Industrial Flavorings



(c) Design Palette for Pickup Trucks

(d) User Design for Laptop Computer Bags

Recently, as preference measurement has moved to web-based questionnaires, e-commerce firms and academic researchers have experimented with a new format in which respondents (or e-commerce customers) select features to configure a preferred product. The best-known e-commerce example is Dell.com (Figure 2a); however, the format is becoming popular on many websites, especially automotive websites (e.g., kbb.com's auto choice advisor, vehix.com, autoweb.com). For marketing research, configurators are known variously as innovation toolkits, design palettes, and user design formats. They have been instrumental in the design of flavorings for International Flavor and Fragrance, Inc (Figure 2b), new truck platforms for General Motors (Figure 2c), and new messenger bags for Timbuk2 (Figure 2d). (Figures from Dahan and Hauser 2002; Thomke and von Hippel 2002, Urban and Hauser 2003; von Hippel 2001). Respondents find configurators easy to use and a natural means to express preferences. However, because configurators focus on the respondents' preferred product, they have only been used to estimate partworth estimates when respondents are asked to repeat the tasks with varying stimuli (e.g., Liechty, Ramaswamy and Cohen 2001).

In this paper, we examine question design for all three forms of preference measurement: metric pairs, stated choices, and configurators. We use formal modeling methods to examine key properties that have been used by the algorithms proposed to date. With these models we seek to understand basic properties in order to help explain why some algorithms perform well and others do not. These models also provide insight on the role of configurators in preference measurement and indicate the strengths and weaknesses of this new form of data collection relative to more-standard methods. The models themselves are drawn from empirical experience and represent, in a stylized manner, the basic properties of question design.

The formal models also provide insight on the relationship of question design to managerial use suggesting, for example, that some measurement formats are best for developing a single product, some for product-line decisions, and some for e-commerce decisions. These insights lead to revised managerial criteria (M_A-error and M_D-error).

The paper is structured as follows:

- We examine first the concept of utility balance as applied to the metric paired-comparison format and demonstrate that this criterion leads to both bias and inefficiency for that format. This analysis also lends insight on recent comparisons between ACA and polyhedral methods (Toubia, et. al. 2003).
- We then examine a related concept for the stated-choice format. The literature often labels this concept "utility balance," but, for the purpose of this paper, we label it "choice balance" to indicate its relationship to stated-choice data. Our analyses indicate that, unlike utility-balance for metric data, choice balance for stated-choice data can lead to improved efficiency under some conditions. In other cases, optimal efficiency requires unbalanced choices. Our analyses both reflect insights from recent algorithmic papers and provide new insights on how choice balance changes as a function of the characteristics of the problem.
- We then examine data that suggests that configurator questions provide more useful information for e-commerce decisions than "tradeoff" questions (Figure 1). We explore this observation with a stylized model to suggest how question formats are best matched to managerial decisions. These insights generalize to M-efficiency.

We define our scope, in part, by what we do not address. Although we draw experience from and seek to provide insight for new algorithms, algorithmic design is not the focus of this paper. Nor do we wish to enter the debate on the relative merits of metric-rating or stated-choice data. For now we leave that debate to other authors such as Elrod, Louviere and Davey (1992) or Moore (2003). We study both data formats (and configurators) and use formal models to gain insights on all three forms of preference questions. We do not address hierarchical models of respondent heterogeneity, but rather focus on each respondent individually. We recognize the practical importance of hierarchical models, but abstract from such estimation to focus on other basic properties. The recent emergence of individual-specific question designs and/or differentiated designs suggest that our insights should extend to such analyses (Johnson, Huber, and Bacon 2003; Sandor and Wedel 2003; Toubia, Hauser, and Simester 2003). But that remains future research.

2. Efficiency in Question Design

Before we begin our analyses, we review briefly the concepts of efficient question design. The concept of question efficiency has evolved to represent the accuracy with which partworths are estimated. Efficiency focuses on the standard errors of the estimates. Let $\vec{p}$ be the vector of partworths to be estimated and $\hat{\vec{p}}$ the vector of estimated partworths. Then, if $\hat{\vec{p}}$ is (approximately) normally distributed with variance Σ, the confidence region for the estimates is an ellipsoid defined by $(\vec{p}-\hat{\vec{p}})'\Sigma^{-1}(\vec{p}-\hat{\vec{p}})$, Greene (1993, p. 190).[1] For most estimation methods, Σ depends upon the questions that the respondent is asked and, hence, an efficient set of questions minimizes the confidence ellipsoid. This is implemented as minimizing a norm of the matrix, Σ. A-errors are based on the trace of Σ, D-errors on the determinant of Σ, and G-errors on the maximum diagonal element. Because the determinant is often the easiest to optimize, most empirical algorithms work with D-efficiency (Kuhfeld, Tobias, and Garratt 1994, p. 547). G-efficiency is rarely used.

Because the determinant of a matrix is the product of its eigenvalues and the trace is the sum of its eigenvalues, A-errors and D-errors are often highly correlated (Kuhfeld, Tobias, and Garratt 1994, p. 547, Arora and Huber 2001, p. 274). In this paper we use both, choosing the cri-

[1] In practice, Σ is unknown and is replaced with its estimated value. For simplicity we assume that Σ is known, which should change neither the insights nor the results of our analyses.

terion which leads to the most transparent insight. In some cases the results are easier to explain if we work directly with Σ^{-1} recognizing $\det(\Sigma) = [\det(\Sigma^{-1})]^{-1}$.[2]

3. Utility Balance for the Metric Paired-Comparison Format

The metric pairs format (Figure 1a) is widely used for commercial conjoint analysis (Green, Krieger, Wind 2001, p. S66; Ter Hofstede, Kim and Wedel 2002, p. 259). Although the format has been used academically and commercially for over 25 years, its most common application is in the adaptive component of ACA (Green, Krieger and Agarwal 1991, Hauser and Shugan 1980, Johnson 1991).[3]

Metric utility balance is at the heart of ACA. With the goal that "a difficult choice provides better information for further refining utility estimates," "ACA presents to the respondent pairs of concepts that are as nearly equal as possible in estimated utility (Orme 1999, p. 2; Sawtooth Software, 2002, p. 11, respectively)." This criterion has appeal. For example, in the study of response latencies, Haaijer, Kamakura, and Wedel (2000, p. 380) suggest that respondents take more time on choice sets that are balanced in utility and, therefore, make less error-prone choices. Green, Krieger and Agarwal (1991, p. 216) quoting a study by Huber and Hansen (1986) ascribe better predictive ability "to greater respondent interest in these difficult-to-judge pairs." Finally, Shugan (1980) provides a theory to suggest that the cost of thinking is inversely proportional to the square of the utility difference. There are clear advantages in terms of respondent interest, attention, and accuracy for metric utility-balanced questions.

However, recent simulation and empirical evidence suggests that alternative criteria provide question designs that lead to more accurate partworth estimates (Toubia, et. al. 2003). Furthermore, adaptation implies endogeneity in the design matrix, which, in general, leads to bias. Thus, we explore whether or not the criterion of metric utility balance has adverse characteristics, such as systematic biases, that offset the advantages in respondent attention. We begin with a stylized model.

---

[2] The trace of a matrix does not necessarily equal the inverse of the trace of the inverse matrix, thus an empirical algorithm that maximizes the trace of Σ^{-1} does not necessarily minimize the trace of Σ. However, the eigenvalues are related, especially when the eigenvalues of Σ are similar in magnitude, as they are in empirical situations.

[3] For example, Dr. Lynd Bacon, AMA's Vice President for Market Research and Executive Vice President of NFO Worldgroup confirms that metric pairs remain one of the most popular forms of conjoint analysis. Sawtooth Software claims that ACA is the most popular form of conjoint analysis in the world, likely has the largest installed base, and, in 2001, accounted for 37% of their sales. Private communications and *Marketing News,* 04/01/02, p. 20.

A Stylized Model

For simplicity we consider products with two features, each of which is specified at two levels. We assume further that the features satisfy preferential independence such that we can write the utility of a profile as an additive sum of the partworths for the features of the profile (Keeney and Raiffa 1976, p. 101-105; Krantz, et. al. 1971). Without loss of generality, we scale the low level of each feature to zero. Let p_1 be the partworth of the high level of feature 1 and let p_2 be the partworth of the high level of feature 2. For these assumptions there are four possible profiles which we write as {0, 0}, {0, 1}, {1, 0}, and {1, 1} to indicate the levels of features 1 and 2, respectively. With this notation we have the true utilities given by:

$$u(0, 0) = 0 \qquad u(1, 0) = p_1 \qquad u(0, 1) = p_2 \qquad u(1, 1) = p_1 + p_2$$

For metric questions in which the respondent provides an interval-scaled response, we model response error as an additive, zero-mean random variable, e. For example, if the respondent is asked to compare {1, 0} to {0, 1}, the answer is given by $a_{12} = p_1 - p_2 + e$. We denote the probability distribution of e with $f(e)$. We denote the estimates of the partworths, estimated from responses to the questions, by $\hat{p}_1$ and $\hat{p}_2$.

Adaptive (Metric) Utility Balance

Without loss of generality, consider the case where $2p_2 > p_1 > p_2 > 0$ such that the off-diagonal question ({0,1} vs. {1,0}) is the most utility-balanced question. We assume that this ordinal relationship is based on prior questions. We label the error associated with the most utility-balanced question as e_{ub} and label errors associated with subsequent questions as either e_1 or e_2. We now consider an algorithm which adapts questions based on utility balance. Adaptive metric utility balance implies the following sequence.

First question:	$\hat{p}_1 - \hat{p}_2 = p_1 - p_2 + e_{ub}$			
Second question:	$\hat{p}_1 = p_1 + e_1$	if	$e_{ub} < p_2 - p_1$	(Case 1)
	$\hat{p}_2 = p_2 + e_2$	if	$e_{ub} \geq p_2 - p_1$	(Case 2)

Suppose that $e_{ub} \geq p_2 - p_1$, then:

$$E[\hat{p}_2] = p_2 + E[e_2] = p_2$$

$$E[\hat{p}_1] = E[\hat{p}_2] + p_1 - p_2 + E[e_{ub} \mid e_{ub} \geq p_2 - p_1] = p_1 + E[e_{ub} \mid e_{ub} \geq p_2 - p_1] > p_1$$

Suppose that $e_{ub} < p_2 - p_1$, then:

$$E[\hat{p}_1] = p_1 + E[e_1] = p_1$$

$$E[\hat{p}_2] = E[\hat{p}_1] + p_2 - p_1 - E[e_{ub} | e_{ub} < p_2 - p_1] = p_2 - E[e_{ub} | e_{ub} < p_2 - p_1] > p_2$$

Thus, under both Case 1 and Case 2, one of the partworths is biased upwards for any zero-mean distribution that has non-zero measure below $p_2 - p_1$.

The intuition underlying this proof is that the second question in the adaptive algorithm depends directly on the random error – a direct version of endogeneity bias (Judge, et. al. 1985, p. 571). We demonstrate how to generalize this result with a pseudo-regression format of $\hat{\vec{p}} = (X'X)^{-1}(X'\vec{a}) = \vec{p} + (X'X)^{-1}X'\vec{e}$, where $\vec{a}$ and $\vec{e}$ are the answer and error vectors and X is the question matrix. Then the second row of X depends on the magnitude of e_{ub} relative to $p_2 - p_1$. Specifically, the first row of X is [1,-1] and its second row can be written as [b, 1-b] where $b = 1$ if errors are small and $b = 0$ if errors are large. The bias, $\vec{B}$, is then given by the following equation where $k = 1$ or 2.

(1) $$E[\vec{B}] = E[(X'X)^{-1}X'e] = E(\begin{bmatrix} (1-b)e_{ub} + e_k \\ -be_{ub} + e_k \end{bmatrix}) = E(\begin{bmatrix} (1-b)e_{ub} \\ -be_{ub} \end{bmatrix})$$

Equation 1 illustrates the phenomenon clearly: $b = 1$ if e_{ub} is small (more negative) and $b = 0$ if e_{ub} is large (more negative), hence either $\hat{p}_1$ or $\hat{p}_2$ is biased upward. These arguments generalize readily to more than two features and more than two questions. The basic insight is that the errors from earlier questions determine the later questions (X) in a systematic way – the essence of endogeneity bias.

Relative Bias

Bias alone does not necessarily impact managerial decisions because partworths are unique only to a positive linear transformation. Thus, we need to establish that the bias is also relative, i.e., the bias depends on the magnitude of the true partworth. Because the bias is proportional to e_{ub} we need to show that $\frac{\partial}{\partial p_1} E[e_{ub} | e_{ub} \geq p_2 - p_1] < 0$ to demonstrate that bias is greater for smaller (true) partworths than it is for larger (true) partworths. We do so by direct differentiation in the Appendix. Thus, we can state the following proposition.

Proposition 1. For the stylized model, on average, metric utility balance biases partworths upward and does so differentially depending upon the true values of the partworths.

When we expand the adaptive questions to a more complex world, in which there are more than two features, the intuition still holds; however, the magnitude of the bias depends upon the specific manner by which metric utility balance is achieved and on the accuracy of the respondent's answers. As an illustration, we built a 16-question simulator for a two-partworth problem in which, after the first question, all subsequent questions were chosen to achieve metric utility balance based on partworths estimated from prior questions. Based on 6,000 regressions each, biases ranged from 1% for relatively large true partworths to 31% for relatively small partworths.[4] To examine the practical significance we cite data from Toubia, et. al. (2003) who simulate ACA question design for a 10-partworth problem. In their data, biases were approximately 6.6% of the magnitudes of the partworths at eight questions when averaged across the conditions in their experimental design. The biases are significant at the 0.05 level.

The Effect of Metric Utility Balance on Efficiency

Although metric utility balance leads to bias, it might be the case that adaptation leads to greater efficiency. For D-errors, minimizing the determinant of Σ is the same as maximizing the determinant of Σ^{-1}. For metric data, $\Sigma^{-1} = X'X$ (for appropriately centered X). From this expression we see two things immediately. First, as Kanninen (2001, p. 217) and Kuhfeld, et. al. (1994, p. 547) note, for a given set of levels, efficiency tends to push features to their extreme values. (These statements assume that the maximum levels are selected to be as large as is reasonable for the empirical problem. Levels which are too extreme risk additional respondent error not explicitly acknowledged by the statistical model used in question design. See Delquie 2003). Second, perfect metric utility balance imposes a linear constraint on the columns of X because $X\vec{p} = \vec{0}$. This constraint induces $X'X$ to be singular. When $X'X$ is singular, the determinant will be zero and D-errors increase without bound. Imperfect metric utility balance leads to

[4] The simulator is available from the website listed in the acknowledgements section of this paper. The simulator is designed to demonstrate the phenomenon; we rely on Toubia, et. al. (2003) for an estimate of the empirical magnitude. The simulator chooses x_1 uniformly on [0,1] and x_2 based on utility balance. The absolute values of the response errors are approximately half those of the dependent measure. The magnitude of the bias changes if the response errors change, but the direction of the bias remains.

det($X'X$) ≈ 0 and extremely large D-errors. A-errors behave similarly. Thus, greater metric utility balance tends to make questions inefficient.[5]

Insights on the Relative Performance of Metric Polyhedral Methods

The analysis of metric utility balance provides insights on the relative performance of metric polyhedral methods. Metric polyhedral methods focus on asking questions to reduce the feasible set of partworths as rapidly as possible by solving an eigenvalue problem related to Sonnevend ellipsoids. These ellipsoids focus questions relative to the "axis" about which there is the most uncertainty. Polyhedral methods do not attempt to impose metric utility balance. In algorithms to date, polyhedral methods are used for (at most) the first n questions where n equals the number of partworths. However, metric polyhedral methods also impose the constraint that the rows of X be orthogonal (Toubia, et. al. 2003, Equation A9). Thus, after n questions, X will be square, non-singular, and orthogonal ($XX' \propto I$ implies $X'X \propto I$).[6] Subject to scaling, this orthogonality relationship minimizes D-error (and A-error). Thus, while the adaptation inherent in metric polyhedral methods leads to endogenous question design, the orthogonality constraint appears to overcome the tendency of endogeneity to increase errors. Toubia, et. al. (2003) report at most a 1% bias for metric polyhedral question selection, significantly less than observed for ACA question selection.

Summary

Metric utility balance may retain respondent interest and encourage respondents to think hard about tradeoffs, however, these benefits come at a price. Metric utility balance leads to bias, relative bias, and lowered efficiency. On the other hand, the inherent orthogonality criterion (rather than utility balance) in metric polyhedral methods enables metric polyhedral question design to achieve the benefits of adaptation with significantly less bias and, apparently, without reduced efficiency.

4. Choice Balance for the Stated-Choice Format

In the stated-choice format (Figure 1b), two profiles that have the same utility will be equally likely to be chosen, thus choice balance is related to utility balance. However, because

---

[5] The simulator cited in the acknowledgements section of this paper also contains an example in which metric utility balance can be imposed. The more closely utilities are balanced, the worse the fit.

[6] Orthogonal rows assume that $XX' = \alpha I$ where α is a proportionality constant. Because X is non-singular, X and X' are invertible, thus $X'XX'X'^{-1} = \alpha X'X'^{-1} \Rightarrow X'X = \alpha I$.

the underlying statistical models are fundamentally different, the implications of choice balance for stated-choice questions are different from the implications of metric utility for metric paired-comparison questions.

Choice balance is an important criterion in many of the algorithms used to customize stated-choice questions. For example, Huber and Zwerina (1996) begin with orthogonal, level-balanced, minimal overlap designs and demonstrate that swapping and relabeling these designs for greater choice balance increases D_p-efficiency. Arora and Huber (2001, p. 275) test these algorithms in a variety of domains with hierarchical Bayes methods commenting that the maximum D_p-efficiency is achieved by sacrificing orthogonality for better choice balance.[7] Orme and Huber (2001, p. 18) further advocate choice balance. Toubia, Hauser, and Simester (2003) use choice-balance to increase accuracy in adaptive choice-based questions. Finally, the concept of choice balance is germane to the principle, used in the computer adaptive testing literature, that a test is most effective when its items are neither too difficult nor too easy, such that the probabilities of correct responses are close to 0.50 (Lord 1970).

To understand better why choice balance improves efficiency while metric utility balance degrades efficiency, we recognize that, for logit estimation of partworths with stated-choice data, Σ^{-1} depends upon the true partworths and, implicitly, on response error. Specifically, McFadden (1974) shows that:

$$\Sigma^{-1} = R\sum_{i=1}^{q}\sum_{j=1}^{J_i}(\vec{x}_{ij} - \sum_{k=1}^{J_i}\vec{x}_{ik}P_{ik})' P_{ij}(\vec{x}_{ij} - \sum_{k=1}^{J_i}\vec{x}_{ik}P_{ik}) \tag{2}$$

where R is the effective number of replicates; J_i is the number of profiles in choice set i; q is the number of choice sets; $\vec{x}_{ij}$ is a row vector describing the j^{th} profile in the i^{th} choice set; and P_{ij} is the probability that the respondent chooses profile j from the i^{th} choice set.

Equation 2 has a more transparent form when all choice sets are binary, in particular:

$$\Sigma^{-1} = R\sum_{i=1}^{q}(\vec{x}_{i1} - \vec{x}_{i2})' P_{i1}(1-P_{i1})(\vec{x}_{i1} - \vec{x}_{i2}) = R\sum_{i=1}^{q}\vec{d}_i{}' P_{i1}(1-P_{i1})\vec{d}_i \tag{3}$$

where $\vec{d}_i$ is the row vector of differences in the features for the i^{th} choice set. In Equation 3 it is clear that choice balance ($P_{i1}\rightarrow ½$) tends to increase a norm of Σ^{-1}. However, the choice prob-

[7] D_p-efficiency recognizes that Σ depends on the true partworths. See Equation 2.

abilities (P_{i1}) are a function of the feature differences ($\vec{d}_i$), thus pure choice balance may not maximize either the determinant (or the trace) of Σ^{-1}.

First recognize that the true choice probabilities in the logit model are functions of the utilities, $\vec{d}_i\vec{p}$. If we hold the utilities constant, we also hold the choice probabilities constant. Subject to this linear utility constraint, we can increase efficiency without bound by increasing the absolute magnitudes of the $\vec{d}_i$'s. This result is similar to that for the metric-pairs format; the experimenter should choose feature differences that are as large as reasonable subject to the linear constraint imposed by $\vec{d}_i\vec{p}$ and subject to any errors (outside the model) introduced by large feature differences (Delquie 2003).

Kanninen (2002) exploits these properties effectively by selecting one continuous feature (e.g. price) as a focal feature with which to manipulate $\vec{d}_i\vec{p}$, while treating all remaining features as discrete by setting them to their maximum or minimum levels. She reports substantial increases in efficiency for both binary and multinomial choice sets (p. 223).

Kanninen uses numerical means to select the optimal levels of the continuous focal feature. We gain further insight with formal analysis. In particular, we examine the optimality conditions for the focal feature, holding the others constant. For greater transparency we use binary questions and the trace of Σ^{-1} rather than the determinant.[8] The trace is given by:

(4) $$trace(\Sigma^{-1}) = \sum_{i=1}^{q}\sum_{k=1}^{K} d_{ik}^2 P_{i1}(1-P_{i1})$$

If we assume, without loss of generality, that the focal feature is the K^{th} feature, then the first-order conditions for the focal feature are:

(5) $$d_{iK} = \left(p_K \sum_{k=1}^{K} d_{ik}^2\right)\left(\frac{P_{i1}-P_{i2}}{2}\right) \text{ for all } i$$

where d_{ik} is the level of the k^{th} feature difference in the i^{th} binary question.

We first rule out the trivial solution of $d_{ik} = 0$ for all *k*. Not only would this imply a choice between two identical alternatives, but the second-order conditions imply a minimum. A slightly more interesting case occurs when perfect choice balance is already achieved by the *K-1*

[8] For a focal feature, the trace and the determinant have related maxima, especially if all but the non-focal feature are set to their minimum or maximum values.

discrete features. We return to this solution later, however, in the focal-feature case it occurs with zero probability (i.e., on a set of zero measure). The more interesting solutions for a continuous focal feature require that d_{iK} be non-zero and, hence, that $P_{i1} \neq P_{i2}$. Thus, when d_{iK} can vary continuously, Equation 5 implies that optimal efficiency requires choice imbalance.

There are two equivalent solutions to Equation 5 corresponding to a right-to-left switch of the two alternatives. That is, the sign of d_{iK} must match the sign of $P_{i1} - P_{i2}$. Thus, without loss of generality, we examine situations were d_{iK} is positive (see Appendix). We plot the marginal impact of d_{iK} on $trace(\Sigma^{-1})$ for illustrative levels of the non-focal feature difference.[9] Following Arora and Huber (2001) and Toubia, Hauser and Simester (2003), we use the magnitude of the partworths as a measure of response accuracy – higher magnitudes indicate higher response accuracy. Figure 3 illustrates that efficiency is maximized for non-zero focal feature differences. Figure 3 also illustrates that the optimal level of the focal feature difference depends upon the level of the non-focal features and upon the magnitude.

Choice balance ($P_{i1} - P_{i2}$) is a non-linear function of the feature differences. Nonetheless, we can identify the systematic impact of both non-focal features and response accuracy. In particular, formal analyses reveal that choice balance increases with greater response accuracy and with greater levels of the non-focal feature differences. We formalize these insights in two propositions. The proofs are in the Appendix.

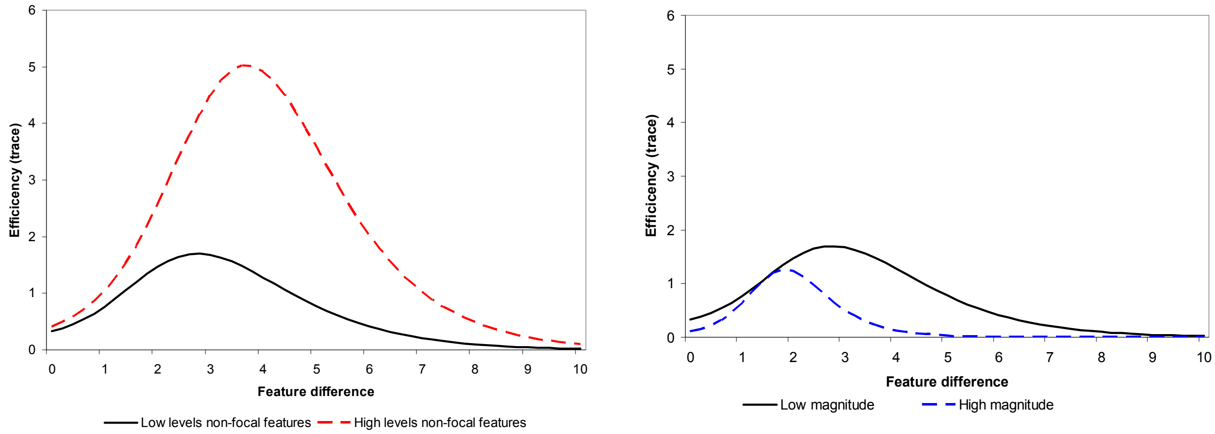
Proposition 2. As response accuracy increases, choice balance increases and the level difference in the focal feature decreases.

Proposition 3. As the levels of the non-focal features increase, choice balance increases.

---

[9] For low magnitudes, the partworths of both feature are set equal to 1.0 in these plots. For high magnitudes they are doubled. The low level of the non-focal feature (absolute value) is 1.5. The high level is 3.0.

Figure 3
Optimal Efficiency and Choice Imbalance



(a) Low vs. high non-focal feature levels (b) Low vs. high magnitude in logit model

Quantal Features

In many applications, all features are binary – they are either present or absent. Examples include four-wheel steering on automobiles, cell-phone holders on laptop bags, and auto focus on cameras. In these cases, the focal-feature method cannot be used for maximizing efficiency. However, discrete-search algorithms are feasible to maximize a norm on Σ^{-1}. Arora and Huber (2001) and Sandor and Wedel (2001) provide two such algorithms. Sandor and Wedel (2002) extend these concepts to a mixed logit model.

An interesting (and common) case occurs when the experimenter limits the number of features that vary, perhaps due to cognitive load (Sawtooth Software 1996, p. 7; Shugan 1980). For a given number of binary features (±1), $\sum_k d_{ik}^2$ is fixed. In this case, it is easy to show that Equation 4 is maximized when choices are as balanced as feasible.[10] This same phenomenon applies when the researcher requires that all features vary.

Insights on the Relative Performance of Choice-based Polyhedral Methods

Like metric polyhedral methods, choice-based polyhedral methods rely upon Sonnevend ellipsoids. However, for stated-choice data, respondents' answers provide inequality constraints rather than equality constraints. Rather than selecting questions vectors parallel to the longest axis of the ellipsoid, choice-based polyhedral methods use the ellipsoid to identify target part-

[10] If there are a maximum number of features that can be non-zero rather than a fixed number, it is possible, although less likely, that a maximum will might occur for less than the maximum number of features. This case can be checked numerically for any empirical problem. Nonetheless, the qualitative insight holds.

worths at the extremes of the feasible region. The choice-based algorithm then solves a utility maximization (knapsack) problem to identify the features of the profiles. The knapsack problem assures that the constraints go (approximately) through the analytic center of the region. This criterion provides approximate a priori choice balance. (By a priori we mean that, given the prior distribution of feasible points, the predicted choices are equal.) In this manner, the algorithm achieves choice balance without requiring utility balance throughout the region. Utility balance only holds for the analytic center of the region – a set of zero measure.

The applications of choice-based polyhedral methods to date have focused on discrete features implemented as binary features. For discrete features, choice balance achieves high efficiency. For continuous features, Proposition 2 suggests algorithmic improvements. Specifically, the knapsack problem might be modified to include *trace*(Σ^{-1}) in the objective function. However, unlike a knapsack solution, which is obtained rapidly, the revised math program may need to rely on new, creative heuristics.

Summary

Choice balance affects the efficiency of stated-choice questions quite differently than metric utility balance affects metric paired-comparison questions. This, despite the fact that choice balance is related to ordinal utility balance. If at least one feature is continuous, then that feature can be used as a focal feature and the most efficient questions require choice imbalance (and non-zero levels of the focal feature). However, we show formally that more accurate responses and higher non-focal features lead to greater choice balance. If features are discrete and the number which can vary is fixed, then choices should be as balanced as feasible. Finally, this analysis helps explain the relative performance of stated-choice polyhedral methods and aggregate customization methods.

5. Configurators

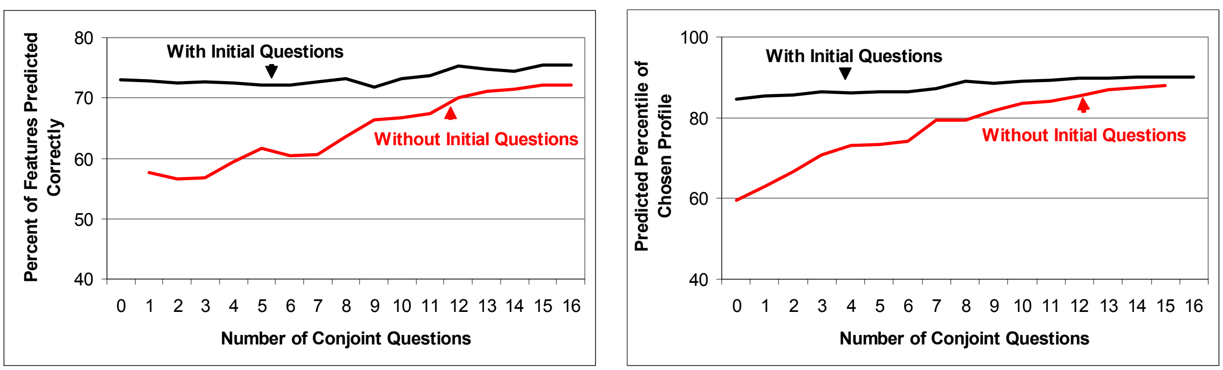
In Sections 3 and 4 we explored two popular formats for preference questions suggesting that, for the metric paired-comparison format, utility balance led to bias, relative bias and inefficiency while, for the choice-based format, choice balance could be efficient for discrete features but not necessarily for continuous focal features. We now explore a relatively new format, configurators, in which respondents focus on the features one at a time and are not asked to compare profiles in the traditional sense. Neither utility balance nor choice balance apply to configurators. Nonetheless, analysis of this new format leads to insights that generalize to all three question formats and helps us understand which format is most efficient in which situations.

In configurators, respondents are typically given a price for each feature level and asked to indicate their preferences for that price/feature combination (review Figure 2). Configurator questions are equivalent to asking the respondent whether the partworths of the features (p_k's) are greater than or less than the stated prices. These data appear quite useful. For example, Liechty, Ramaswamy and Cohen (2001) use ten repeated configurator measures per respondent to estimate successfully hierarchical Bayes probit models.

Stylized Fact: Warm-up Questions do Surprisingly Well in Predicting Feature Choice

Figure 4 presents an empirical example. In data collected by Toubia, et. al. (2003), respondents were shown five laptop bags worth approximately $100 and allowed to choose among those bags. Various conjoint methods based on metric paired-comparison questions collected prior to this task predicted these profile choices quite well.

Figure 4
Warm-up Questions are Effective Predictors of Configurator Tasks



(a) Ability to predict feature choice (b) Ability to predict preferred profile

Before completing the metric paired-comparison task, in order to familiarize respondents with the features, respondents answered warm-up questions. They were asked to indicate the preferred level of each feature. Such warm-up questions are common. We re-analyzed Toubia, et. al.'s data in two ways. One conjoint estimation method did not use the warm-up questions. The other conjoint estimation method included monotonic constraints based on the warm-up questions.[11] One month after the initial conjoint-analysis questionnaire, the respondents were

[11] In the method labeled "with initial questions," the warm-up questions where used to determine which of two feature levels were preferred (e.g., red or black). In the method labeled "without initial questions," such ordering data

recontacted and given a chance to customize their chosen bag with a configurator. Figure 4 plots the ability of the conjoint analyses to predict configurator choice as a function of the number of metric paired-comparison questions that were answered.

For both predictive criteria, the ability to predict configurator choice is surprising flat when initial warm-up questions are included. (Predictions are less flat for forecasts of the choice of five laptop bags, Toubia, et. al. 2003, Table 6.) For configurator predictions, the nine warm-up questions appear to provide more usable information than the sixteen paired-comparison questions. This empirical observation is related to observations by Evgeniou, Boussios, and Zacharia (2003) whose simulations suggest the importance of ordinal constraints. While, at first, Figure 4 appears to be counter-intuitive, it becomes more intuitive when we realize that the warm-up questions are similar to configurator choice and the paired-comparison questions are similar to the choice among five (Pareto) profiles. We formalize this intuition with our stylized model.

Stylized Model to Explain the Efficiency of Warm-up Questions

For transparency assume metric warm-up questions. This enables us to compare the warm-up questions to paired-comparison questions without confounding a metric-vs.-ordinal characterization. We consider the same three questions used previously, rewriting e_{ub} as e_t for mnemonic exposition of paired-comparison tradeoff questions. Without loss of generality we incorporate feature prices into the definition of each feature. We assume $p_1 > p_2 > 0$, but need not assume $2p_2 > p_1$ as we did before. The three questions are:

(Q1) $$\hat{p}_1 = p_1 + e_1$$

(Q2) $$\hat{p}_2 = p_2 + e_2$$

(Q3) $$\hat{p}_1 - \hat{p}_2 = p_1 - p_2 + e_t$$

We now consider stylized estimates obtained from two questions. In our stylized world of two features at two levels, Q1 and Q2 can be either metric warm-up questions or metric paired-comparison questions. Q3 can only be a metric paired-comparison question. Assuming all questions are unbiased, we examine which combination of questions provides the greatest precision (efficiency): Q1+ Q2, Q1+Q3, or Q2+Q3.

---

were not available to the estimation algorithm. In this particular plot, for consistency, we use the analytic-center method of Toubia, et. al., however, the basic insights are not limited to this estimation method.

Consider feature 1. Because feature price is incorporated in the feature, if $p_k \geq 0$, then feature choice is correctly predicted if $\hat{p}_k$ is non-negative. (We use similar reasoning for feature 2.) Simple calculations (see Appendix) reveal that the most accurate combination of two questions is Q1+Q2 if and only if Prob[$e_k \geq -p_k$] > Prob[$e_k + e_t \geq -p_k$]. Since $-p_k$ is a negative number, these conditions hold for most reasonable density functions. For example, in the Appendix we demonstrate that, if the errors are independent and identically distributed normal random variables, then this condition holds.

Proposition 4. For zero-mean normally distributed (i.i.d.) errors, the warm-up questions (Q1+Q2) provide more accurate estimates of feature choice than do pairs of questions that include a metric paired-comparison question (Q3).

Intuitively, metric warm-up questions parse the errors more effectively than tradeoff questions when the focus is on feature choice rather than holistic products. The questions focus precision where it is needed for the managerial decisions. We provide a more formal generalization of these concepts in the next section. The intuition does not depend upon the metric properties of the warm-up questions and applies equally well to ordinal questions. We invite the reader to extend Proposition 4 to ordinal questions using the M-efficiency framework that is explored in the next section.

6. Generalizations of Question Focus: M-Efficiency

The intuition drawn from Proposition 4 suggests that it is best to match preference questions carefully to the managerial task. For example, metric warm-up questions are similar to configurator questions and may be the most efficient for predicting feature choice. On the other hand, tradeoff questions are similar to the choice of a single product from a Pareto set. Estimating partworths from tradeoff questions may be best if the managerial goal is to predict consumer choice when a single new product is launched into an existing market. We now examine this more general question. We begin with the stylized model for two questions and then suggest a generalization to an arbitrary number of questions.

Stylized Analysis: Matching Question Selection to Managerial Focus

With metric questions (and preferential independence) we can ask one more independent question, which we call a "range" question.[12] We then consider three stylized product-development decisions.

(Q4) $$\hat{p}_1 + \hat{p}_2 = p_1 + p_2 + e_r$$

Launch a single product

The product-development manager has one shot at the market and has to choose the single best product to launch, perhaps because of shelf-space or other limitations. The manager is most interested in the difficult tradeoffs, especially if the two features are substitutes. The manager seeks precision on $p_1 - p_2$.

Launch a product line (platform decision)

The product-development manager is working on a platform from which to launch multiple products, perhaps because of synergies in production. The manager is most interested in the range of products to offer (the breadth of the platform). The manager seeks precision on $p_1 + p_2$.

Launch a configurator-based e-commerce website

The product-development manager plans to sell products with a web-based configurator, perhaps because of flexible manufacturing capability. However, there is a cost to maintaining a broad inventory and there is a cognitive load on the consumer based on the complexity of the configurator. The manager seeks precision on p_1 and on p_2.

For these stylized decisions we expect intuitively that we gain the most precision by matching the questions to the managerial decisions. Table 1 illustrates this intuition. For decisions on a single product launch, we gain the greatest precision by including tradeoff questions (Q3); for decisions on product lines we gain the greatest precision by including range questions (Q4), and for decisions on e-commerce we gain the greatest precision by focusing on feature-specific questions (Q1 and Q2). These analytical results, which match question format to managerial need (for metric data), are comparable in spirit to the empirical results of Marshall and Bradlow (2002) who examine congruency between the formats of the profile and validation data (metric vs. ranking vs. choice) in hybrid conjoint analysis. Marshall and Bradlow's analyses suggest that profile data require less augmentation with self-explicated data when they are congruent with the validation data.

[12] Such questions are feasible with a metric format. They would be eliminated as dominated alternatives in an ordinal format such as stated-choice questions.

Table 1
Error Variances Due to Alternative Question Selection

	Tradeoff Questions	Range Questions	Feature Focus
Single-product launch	σ^2	$3\sigma^2$	$2\sigma^2$
Product-line decisions	$3\sigma^2$	σ^2	$2\sigma^2$
e-Commerce website	$\frac{3}{2}\sigma^2$	$\frac{3}{2}\sigma^2$	σ^2

M-Efficiency

The insights from Table 1 are simple, yet powerful, and generalize readily. In particular, the stylized managerial decisions seek precision on $M\vec{p}$ where M is a linear transform of $\vec{p}$. For example, if the manager is designing a configurator and wants precision on the valuation of features relative to price, then the manager wants to concentrate precision not on the partworths per se, but on specific combinations of the partworths. We indicate this managerial focus with a matrix, M_c, where the last column corresponds to the price partworth. This focus matrix provides criteria, analogous to D-errors and A-errors, by which we can design questions.

$$M_c = \begin{bmatrix} 1 & 0 & \ldots & -1 \\ 0 & 1 & \ldots & -1 \\ & & \ldots & \\ 0 & \ldots & 1 & -1 \end{bmatrix}$$

M-Efficiency for Metric Data

We begin with metric data to illustrate the criteria. For metric data, the standard error of the estimate of $M\vec{p}$ is proportional to $\Sigma_M = M(X'X)^{-1}M'$ (Judge, et. al. 1985, p. 57). By defining efficiency based on Σ_M rather than Σ, we focus precision on the partworths that matter most to the manager's decision. We call this focus managerial efficiency (M-efficiency) and define M-errors as follows for suitably scaled X matrices.

$$\text{M}_\text{D}\text{-error} = q \det(M(X'X)^{-1}M')^{1/n}$$

$$\text{M}_\text{A}\text{-error} = q\,trace(M(X'X)^{-1}M')/n$$

The managerial focus affects D-errors and A-errors differently. M_D-error is defined by the determinant of the modified covariance matrix and, hence, focuses on the geometric mean of the estimation variances (eigenvalues). All else equal, the geometric mean is minimized if the estimation variances are equal in magnitude. On the other hand, M_A-error is defined by the trace of the covariance matrix and, hence, focuses on the arithmetic mean of the estimation variances. Thus, if the manager wants differential focus on some combinations of partworths and is willing to accept less precision on other combinations, then M_A-errors might be a better metric than M_D-errors.

To illustrate M-efficiency, we provide an example in which we accept higher A- and D-errors in order to reduce M_A- and M_D-errors. Consider an orthogonal array, *X*. For this design, $X'X = 12I_6$, where I_6 is a 6x6 identify matrix. The orthogonal array minimizes both D-errors and A-errors. It does not necessarily minimize either M_D- or M_A-errors as we illustrate with the five-feature configurator matrix, M_c.

$$M_c = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

$$X = \begin{bmatrix} +1 & -1 & -1 & -1 & -1 & -1 \\ -1 & +1 & -1 & +1 & +1 & +1 \\ +1 & +1 & +1 & -1 & +1 & +1 \\ -1 & -1 & +1 & +1 & -1 & +1 \\ +1 & +1 & -1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 & -1 & +1 \\ -1 & -1 & -1 & -1 & +1 & -1 \\ -1 & +1 & -1 & -1 & -1 & +1 \\ -1 & +1 & +1 & -1 & -1 & -1 \\ +1 & +1 & +1 & +1 & -1 & -1 \\ -1 & -1 & +1 & +1 & +1 & -1 \end{bmatrix}$$

To increase M-efficiency, we perturb $X'_M X_M$ closer to $M'_c M_c$ while maintaining full rank. We choose X_M such that $X'_M X_M = \alpha[\beta I_6 + (1-\beta)M'_c M_c]$ where α and β are scalars, $\beta \in [0,1]$, and α is chosen to maintain constant question "magnitudes."[13] By construction, X_M is no longer an orthogonal array, hence both A-errors and D-errors will increase. However, M_A- and M_D-errors decrease because X_M is more focused on the managerial problem. For example, with $\beta = 4/5$, M_A-errors decrease by over 20%; M_D-errors decrease, but not as dramatically. Larger β's turn out to be better at reducing M_D-errors, but less effective at reducing M_A-errors.

Decreases in M_A-efficiency are also possible for square *M* matrices. However, we cannot improve M_D-errors for square *M* matrices, because, for square *M,* the question design that minimizes D-error will also minimize M_D-error. This is true by the properties of the determinant: $\det(\Sigma_M) = \det(M(X'X)^{-1}M') = \det(M)\det((X'X)^{-1})\det(M') = \det(M')\det(M)\det((X'X)^{-1}) = \det(M'M)\cdot\det((X'X)^{-1})$. As an illustration, suppose the manager wants to focus on configurator prices (as in M_c), but also wants to focus on the highest price posted in the configurator. To capture this focus, we make *M* square by adding another row to M_c, [1 1 1 1 1 -1]. To decrease M_A-error we unbalance *X* to make price more salient in the question design. For example, a 40% unbalancing decreases M_A-errors by 5% with only a slight increase in M_D-errors (<1%).

We chose these examples to be simple, yet illustrative. We can create larger examples where the decreases in M-errors are much larger. Our perturbation algorithms are simple; more complex algorithms might reduce M-errors further. M-efficiency appears to be an interesting and relevant criterion on which to focus future algorithmic development.

M-Efficiency for Stated-Choice Questions

The same arguments apply to stated-choice questions. The M-efficient criteria become either the determinant (M_D-error) or the trace (M_A-error) of $M(Z'PZ)^{-1}M'$ where *Z* is a probability-balanced *X* matrix and *P* is a diagonal matrix of choice probabilities. See Arora and Huber (2001, p. 274) for notation. As in Section 4, there will be a tradeoff between choice balance and the levels of focal features. For M-errors, that tradeoff will be defined on combinations of the features rather than the features themselves.

[13] Recall that efficiency criteria tend to push feature levels to the extremes of the feasible space. Thus, to compare different designs we must attempt to keep the "magnitudes" of the features constant. In this example, we implement that constraint by keeping the trace of $X_M'X_M$ equal to the trace of $X'X$. Our goal is to provide an example where M-efficiency matters. We can also create examples for other constraints – the details would be different, but the basic insights remain.

Summary of Sections 5 and 6

Configurator questions, based on analogies to e-commerce websites, are a relatively recent form of preference questions. Empirical data suggest that such questions are surprisingly good at predicting feature choice, in large part because they focus precision on managerial issues that are important for e-commerce and mass customization. This insight generalizes. Intuitively, precision (efficiency) is greatest when the question format is matched to the managerial decision. M_A-errors (and M_D-errors) provide criteria by which match format to managerial relevance. These criteria can be used to modify question-design algorithms.

There remain the challenges of new algorithmic development. Existing algorithms have been optimized and tested for D- and D_p-efficiency. However, Figure 4, Table 1, and the concept of M-errors have the potential to refocus these algorithms effectively.

7. Summary and Future Research

In the last few years great strides have been made in developing algorithms to select questions for both metric paired-comparison preference questions and for stated-choice questions. Algorithms have been developed for a wide range of estimation methods including regression, logit, probit, mixed logit, and hierarchical Bayes (both metric and choice-based). In some cases the algorithms focus on static designs and in other cases on adaptive designs. This existing research has led to preference questions that are more efficient and more accurate as evidenced by both simulation and empirical experiments.

Summary

In this paper, we use stylized models to gain insight on the properties of algorithms and the criteria they optimize. The models abstract from the complexity of empirical estimation to examine properties transparently. In key cases, the stylized models provide insight on empirical facts. While some of the following insights formalize that which was known from prior algorithmic research, many insights are new and provide a basis for further development:

- Metric utility balance leads to relative biases in partworth estimates.
- Metric utility balance is inefficient.
- Choice balance, while related to utility balance, affects stated-choice questions differently than utility balance affects metric questions.
- When features are discrete and the number of features which can vary is fixed, choice balance is a component of optimal efficiency.

- Otherwise, optimal efficiency implies non-zero choice balance.
- As response accuracy and/or the levels of non-focal features increase, optimal efficiency implies greater choice balance.
- Warm-up (configurator) questions are surprisingly accurate for predicting feature choice because they focus precision on the feature-choice decision.
- This concept generalizes. Tradeoff, range, or configurator questions provide greater efficiency when they are matched to the appropriate managerial problem.
- M-errors provide criteria by which both metric and stated-choice questions can be focused on the managerial decisions that will be made based on the data.
- Simple algorithms (perturbation and/or unbalancing) can decrease M-errors effectively.

The formal models also provide a theory with which to understand the accuracy of recently proposed polyhedral methods. Specifically,

- <u>Metric</u> polyhedral methods appear to avoid the bias and inefficiency of utility balance by maintaining orthogonality while focusing questions where partworths estimates are less precise.
- To date, stated-choice polyhedral methods have been applied for discrete features. Hence, their use of approximate choice balance makes them close to efficient.
- However, stated-choice polyhedral methods for continuous features might be improved with algorithms that seek non-zero choice balance.

Future Research

This paper, with its focus on stylized models, takes a different tack than recent algorithmic papers. The two tacks are complementary. Our stylized models can be extended to explain simulation results in Andrews, Ainslie and Currim (2002), empirical results in Moore (2003), or to explore the efficient use of self-explicated questions (Ter Hofstede, Kim and Wedel 2002). On the other hand, insights on utility balance, choice balance, and M-efficiency can be used to improve algorithms for both aggregate customization and polyhedral methods. Finally, interest in configurator-like questions is growing as more marketing research moves to the web and as e-commerce managers demand marketing research that matches the decisions they make daily. We have only begun to understand the nature of these questions.

References

Andrews, Rick L., Andrew Ainslie, and Imran S. Currim (2002), "An Empirical Comparison of Logit Choice Models with Discrete versus Continuous Representations of Heterogeneity," *Journal of Marketing Research*, 39, (November), 479-487

Arora, Neeraj and Joel Huber (2001), "Improving Parameter Estimates and Model Prediction by Aggregate Customization in Choice Experiments," *Journal of Consumer Research*, 28, (September), 273-283.

Dahan, Ely and John R. Hauser (2002), "The Virtual Customer," *Journal of Product Innovation Management*, 19, 5, (September), 332-354..

Delquie, Philippe (2003), "Optimal Conflict in Preference Assessment," *Management Science*, 49, 1, (January), 102-115.

Evgeniou, Theodoros, Constantinos Boussios, and Giorgos Zacharia (2003), "Generalized Robust Conjoint Estimation," Working Paper, (Fontainebleau, France: INSEAD), May.

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

Green, Paul E, Abba Krieger, and Manoj K. Agarwal (1991), "Adaptive Conjoint Analysis: Some Caveats and Suggestions," *Journal of Marketing Research*, pp. 215-222.

______, ______, and Yoram Wind (2001), "Thirty Years of Conjoint Analysis: Reflections and Prospects," *Interfaces,* 31, 3, Part 2, (May-June), S56-S73.

Greene, William H. (1993), *Econometric Analysis, 2E*, (Englewood Cliffs, NJ: Prentice-Hall, Inc.)

Haaijer, Rinus, Wagner Kamakura, and Michel Wedel (2000), "Response Latencies in the Analysis of Conjoint Choice Experiments," *Journal of Marketing Research* 37, (August), 376-382.

Hauser, John R. and Steven M. Shugan (1980), "Intensity Measures of Consumer Preference," *Operations Research*, 28, 2, (March-April), 278-320.

Huber, Joel and David Hansen (1986), "Testing the Impact of Dimensional Complexity and Affective Differences of Paired Concepts in Adaptive Conjoint Analysis," *Advances in Consumer Research*, 14, Melanie Wallendorf and Paul Anderson, eds., Provo, UT: Associations of Consumer Research, 159-163.

______ and Klaus Zwerina (1996), "The Importance of Utility Balance in Efficient Choice Designs," *Journal of Marketing Research*, 33, (August), 307-317.

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

______, Joel Huber, and Lynd Bacon, "Adaptive Choice Based Conjoint," *Sawtooth Software Proceedings*, 2003.

Kanninen (2002), "Optimal Design for Multinomial Choice Experiments," *Journal of Marketing Research*, 39, (May), 214-227.

Kuhfeld, Warren F. , Randall D. Tobias, and Mark Garratt (1994), "Efficient Experimental Design with Marketing Research Applications," *Journal of Marketing Research*, 31, 4, (November), 545-557.

Lord, Frederic M. (1970), "Some Test Theory for Tailored Testing," Wayne H. Holzman (Ed.), *Computer Assisted Instruction, Testing, and Guidance*, (New York, NY: Harper and Row).

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

______, Joel Huber, and Lynd Bacon (2003), "Adaptive Choice Based Conjoint," *Proceedings of the Sawtooth Software Conference*, April.

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

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

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

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

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

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

Marshall, Pablo, and Eric T. Bradlow (2002), "A Unified Approach to Conjoint Analysis Models," *Journal of the American Statistical Association*, 97, 459 (September), 674-682.

McFadden, Daniel (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," *Frontiers in Econometrics*, P. Zarembka, ed., (New York: Academic Press), 105-142.

Moore, William L. (2003), "A Cross-Validity Comparison of Conjoint Analysis and Choice Models," Working Paper, (Salt Lake City, Utah: University of Utah), February.

Orme, Bryan (1999), "ACA, CBC, or Both: Effective Strategies for Conjoint Research," Working Paper, Sawtooth Software, Sequim, WA.

______ and Joel Huber (2000), "Improving the Value of Conjoint Simulations," *Marketing Research*, 12, 4, (Winter), 12-20.

Sandor, Zsolt and Michel Wedel (2001), "Designing Conjoint Choice Experiments Using Managers' Prior Beliefs," *Journal of Marketing Research*, 38, 4, (November), 430-444.

______ and ______ (2002), "Profile Construction in Experimental Choice Designs for Mixed Logit Models," *Marketing Science*, 21, 4, (Fall), 455-475.

______ and ______ (2003), "Differentiated Bayesian Conjoint Choice Designs," working paper, University of Michigan Business School.

Sawtooth Software (2002), "ACA 5.0 Technical Paper," Sawtooth Software Technical Paper Series, (Sequim, WA: Sawtooth Software, Inc.)

Shugan, Steven M. (1980), "The Cost of Thinking," *Journal of Consumer Research*, 7, 2, (September), 99-111.

Ter Hofstede, Frenkel, Youngchan Kim, and Michel Wedel (2002), "Bayesian Prediction in Hybrid Conjoint Analysis," *Journal of Marketing Research*, 39, (May), 253-261.

Thomke, Stefan and Eric von Hippel (2002), "Customers as Innovators: A New Way to Create Value," *Harvard Business Review*, (April), 74-81.

Toubia, Olivier, John R. Hauser, and Duncan I. Simester (2003), "Polyhedral Methods for Adaptive Choice-based Conjoint Analysis," forthcoming, *Journal of Marketing Research*.

______, Duncan Simester, John R. Hauser, and Ely Dahan (2003), "Fast Polyhedral Adaptive Conjoint Estimation," forthcoming, *Marketing Science*.

Urban, Glen L. and John R. Hauser (2003), "'Listening In' to Find and Explore New Combinations of Customer Needs," forthcoming, *Journal of Marketing*.

von Hippel, Eric (2001b), "Perspective: User Toolkits for Innovation," *Journal of Product Innovation Management*, 18, 247-257.

Appendix

Relative Bias Due to Metric Utility Balance

Proposition 1. For the stylized model, on average, metric utility balance biases partworths upward and does so differentially depending upon the true values of the partworths.

Proof. In the text we have already shown that one of the partworths is biased upwards for any zero-mean distribution that has non-zero measure below $p_2 - p_1$. We must now demonstrate that $\frac{\partial}{\partial p_1} E[e_{ub} \,|\, e_{ub} \geq p_2 - p_1] > 0$ for $p_1 - p_2 \geq 0$ with density functions which have non-zero density below $p_2 - p_1$. We differentiate the integral to obtain:

$$\frac{\partial}{\partial p_1} E[e_{ub} \,|\, e_{ub} \geq p_2 - p_1] = \frac{\partial}{\partial p_1}\left[\int_{p_2-p_1}^{\infty} e_{ub} f(e_{ub})\, de_h \Big/ \int_{p_2-p_1}^{\infty} f(e_{ub}) de_{ub}\right]$$

$$= \frac{\left(\int_{p_2-p_1}^{\infty} f(e_{ub})\, de_{ub}\right)(p_2 - p_1) f(p_2 - p_1) - \left(\int_{p_2-p_1}^{\infty} e_{ub} f(e_{ub})\, de_{ub}\right) f(p_2 - p_1)}{\left(\int_{p_2-p_1}^{\infty} f(e_{ub})\, de_{ub}\right)^2} =$$

$$\frac{-[1 - F(p_2 - p_1)](p_1 - p_2) f(p_2 - p_1) - E[e_{ub} \,|\, e_{ub} \geq p_2 - p_1][1 - F(p_2 - p_1)] f(p_2 - p_1)}{[1 - F(p_2 - p_1)]^2} < 0$$

where the last step recognizes that both terms in the numerator are negative whenever $f(e_{ub})$ has density below $p_2 - p_1$ and $p_1 - p_2 > 0$. When $p_1 = p_2$, the second term is still negative.

Choice Balance vs. Magnitude and Non-focal Features

From Equation 3 in the text we know that $trace(\Sigma^{-1}) = \sum_{i=1}^{q}\sum_{k=1}^{K} d_{ik}^2 P_{i1}(1 - P_{i1})$. Because this function is separable in *i*, we focus on a single *i* and drop the *i* subscript. Rewriting P_i in terms of the partworths, $\vec{p}$, and allowing *m* to scale the magnitude of the partworths, we obtain for a given *i*, $T \equiv trace_i(\Sigma^{-1}) = \sum_{k=1}^{K} d_k^2 \Big/ (e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}} + 2)$. We assume, without loss of generality, that $p_K > 0$ and $y_{K-1} \equiv \sum_{k=1}^{K-1} p_k d_k \leq 0$. Define $s_{K-1} \equiv \sum_{k=1}^{K-1} d_k^2$ and let d_K^* be the optimal d_K.

Lemma 1. For $p_K > 0$ and $y_{K-1} \leq 0$, $d_K^* \geq 0$.

Proof. Assume $d_K^* < 0$. If $a^* = p_K d_K^* + y_{K-1} < 0$, then $trace^* = (d_K^{*2} + s_{K-1})/((e^{ma^*} + e^{-ma^*} + 2)$. Consider $d_K^{**} = (-a^* - y_{K-1})/p_K > 0$ such that $a^{**} = -a^* > 0$. This assures that the denominator of the trace stays the same. Now $\left|d_K^{**}\right| > \left|(-a^* - y_{K-1})/p_K\right| > \left|(a^* - y_{K-1})/p_K\right| = \left|d_K^*\right|$ because both a^* and y_{K-1} are of the same sign. Thus, the numerator of the trace is larger and the denominator is unchanged and we have the result by contraction. In the special case of $y_{K-1} = 0$, $trace(d_K) = trace(-d_K)$, hence we can also restrict ourselves to $d_K \geq 0$.

Lemma 2. For $p_K > 0$, $y_{K-1} < 0$, then the trace has no minimum in d_K on (0, ∞).

Proof. $T'(d_K) = [2d_K(e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}} + 2) - \sum_k d_k^2 m p_K (e^{m\vec{d}\vec{p}} - e^{-m\vec{d}\vec{p}})]/(e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}} + 2)^2 \equiv h(d_K)/b(d_K)$. Then, $T'(d_K^*) = 0 \Rightarrow h(d_K^*) = 0$ and $T''(d_K^*) = h'(d_K^*)/b(d_K^*)$. $T''(d_K^*)$ will have the same sign as $h'(d_K^*) = 2(e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}} + 2) + 2d_K m p_K (e^{m\vec{d}\vec{p}} - e^{-m\vec{d}\vec{p}}) - 2d_K m p_K (e^{m\vec{d}\vec{p}} - e^{-m\vec{d}\vec{p}}) - \sum_k d_k^2 m^2 p_K^2 (e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}})$. Cancel the second and third terms. Using the FOC gives $h'(d_K^*) = \sum_k d_k^2 m p_K (e^{m\vec{d}\vec{p}} - e^{-m\vec{d}\vec{p}})/d_K - \sum_k d_k^2 m^2 p_K^2 (e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}})$. Thus, $h'(d_K^*)$ and $H \equiv (e^{m\vec{d}\vec{p}} - e^{-m\vec{d}\vec{p}}) - m d_K p_K (e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}})$ have the same sign. Because $d_K^* \geq 0$ by assumption, the FOC imply that $e^{m\vec{d}\vec{p}} - e^{-m\vec{d}\vec{p}} \geq 0$, hence $m\vec{d}\vec{p} > 0$. If $m d_K p_K \geq 1$, then H < 0 and so is $T''(d_K^*)$. If $m d_K p_K < 1$, then $m\vec{d}\vec{p} < 1$ because $y_{K-1} < 0$. Thus, $e^0 < e^{m\vec{d}\vec{p}} < e^1$, hence $H < e - 1 - 2m d_K p_K$. Thus, $H < 0$ if $m d_K p_K > (e-1)/2$. Call $h_0 = (e-1)/2$. If $m d_K p_K < h_0$, then $0 < m\vec{d}\vec{p} < h_0$ and $H < e^{h_0} - 1 - 2m d_K p_K$, which is negative if $m d_K p_K > h_1 > (e^{h_0} - 1)/2$. For all *h* in (0,1], we have $(e^h - 1)/2 > h$, so by recursion we show that $H < 0$ if $m d_K p_K > h_\ell$ and h_ℓ converging to zero.

Proposition 2. As response accuracy increases, choice balance increases and the level difference in the focal feature decreases.

Proof. Assume $p_K > 0$ *and* $y_{K-1} \leq 0$ without loss of generality and rewrite $T = u(d_K, m)/v(d_K, m)$ where $u(d_K, m) = \sum_k d_k^2$, $v(d_K, m) = (e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}} + 2)$. For $m_o > 0$, $d_K^*(m_o)$ satisfies $f \equiv u'/u = v'/v \equiv g$. The numerator, *u*, does not depend upon *m*. Taking derivatives yields $v'/v = m p_K (e^{m\vec{d}\vec{p}} - e^{-m\vec{d}\vec{p}})/(e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}} + 2)$ and $\partial(v'/v)/\partial m =$

$$\frac{[p_K(e^{m\vec{d}\vec{p}} - e^{-m\vec{d}\vec{p}}) + m p_K \vec{d}\vec{p}(e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}})][e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}} + 2] - m p_K \vec{d}\vec{p}(e^{m\vec{d}\vec{p}} - e^{-m\vec{d}\vec{p}})^2}{(e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}} + 2)^2}.$$

We rearrange the numerator to obtain the following expression: $p_K(e^{m\vec{d}\vec{p}} - e^{-m\vec{d}\vec{p}})(e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}} + 2) + 2m p_K \vec{d}\vec{p}(e^{m\vec{d}\vec{p}} + e^{-m\vec{d}\vec{p}}) + 4m p_K \vec{d}\vec{p}$, which is positive because $m\vec{d}\vec{p} > 0$ and $e^{m\vec{d}\vec{p}} - e^{-m\vec{d}\vec{p}} \geq 0$ as proven in Lemma 2. Hence, v'/v is increasing in *m*. Thus, $g(d_K^*(m_o), m_1) > g(d_K^*(m_o), m_o)$ for $m_1 > m_o$, which implies $f(d_K^*(m_o), m_1) = f(d_K^*(m_o), m_o) = g(d_K^*(m_o), m_o) < g(d_K^*(m_o), m_1)$. Hence, $T'(d_K^*(m_o), m_1) < 0$.

We have shown that at $m = m_1$, the derivative to *T* with respect to d_K is negative at

$d_K^*(m_o)$. By Lemma 2, it is non-positive for all $d_K > d_K^*(m_o)$. Thus, $d_K^*(m_1) < d_K^*(m_o)$ which proves the second part of the proposition. Because $y_{K-1} + d_K^*(m_o)p_K > 0$, $y_{K-1} + d_K^*(m_1)p_K > 0$, and $y_{K-1} < 0$, we have $0 < y_{K-1} + d_K^*(m_1)p_K < y_{K-1} + d_K^*(m_o)p_K$, which proves the first part of the proposition.

Proposition 3. As the levels of the non-focal features increase, choice balance increases.

Proof. We now consider changes in the non-focal features. In this case, *m* scales the non-focal features but does not scale d_K. Thus, $T = (m^2 y_{K-1} + d_K^2)\big/(e^{my_{K-1}+d_K p_K} + e^{-(my_{K-1}+d_K p_K)} + 2)$. We again rewrite $T = u(d_K, m)/v(d_K, m)$. Caution, the definitions of *u* and *v* (and *f* and *g*) are <u>within</u> the proofs and vary between the two propositions. Lemmas 1 and 2 still hold because the proofs follow the same principles. Thus, w.l.o.g., we have $p_K > 0$, $y_{K-1} \le 0$ which implies that $d_K^* \ge 0$ and that $my_{K-1} + d_K^* p_K > 0$. This is true for any $m > 0$.

Let $m_o > 0$, then the FOC imply $f(d_K^*, m_o) \equiv v'/v = u'/u \equiv g(d_K^*, m_o)$. Let $m_1 \equiv m_o + dm_o$ and consider a $d_K = d_K^*(m_o) + dd_K$ such that utility is unchanged: $a^*(m_o) \equiv m_o y_{K-1} + d_K^*(m_o)p_K$ $=(m_o + dm_o)y_{K-1} + (d_K^*(m_o) + dd_K)p_K$. Solving we obtain $dd_K = -dm_o y_{K-1}/p_K$. For reference call this Expression 1. Differentiating w.r.t. d_K gives $f(d_K^*, m_o) = p_K(e^{a^*} - e^{-a^*})/(e^{a^*} + e^{a^*} + 2)$. Because a^* is unchanged we have $f(d_K^*(m_o) + dd_K, m_o + dm_o) = f(d_K^*(m_o), m_o)$.

Differentiating *u* w.r.t. d_K gives: $g(d_K^*, m_o) = 2d_K^*/(m_o^2 s_{K-1} + d_K^{*2})$. Further differentiation gives $\partial g(d_K^*, m_o)/\partial d_K = (2m_o^2 s_{K-1} - 2d_K^{*2})/(m^2 s_{K-1}^2 + d_K^{*2})^2$ and $\partial g(d_K^*, m_o)/\partial m =$ $-4d_K^* m s_{K-1}/(m^2 s_{K-1} + d_K^{*2})^2$. Consider $dm_o > 0$. Using Taylor's Theorem, ignoring higher order terms because dm_o and, hence, dd_K are differentials: $g(d_K^* + dd_K, m_o + dm_o) \approx g(d_K^*, m_o) +$ $(2m_o^2 s_{K-1} - 2d_K^{*2})/(m_o^2 s_{K-1}^2 + d_K^{*2})^2 dd_k - 4d_K^* m_o s_{K-1}/(m_o^2 s_{K-1} + d_K^{*2})^2 dm_o$. Expression 1 implies $G_o \equiv g(d_K^* + dd_K, m_o + dm_o) - g(d_K^*, m_o)$ is of the same sign as $G_1 \equiv -4d_K^*(m_o)m_o s_{K-1}$ $-(2m_o^2 s_{K-1} - 2d_K^{*2}(m_o))y_{K-1}/p_K$. Because $y_{K-1} < 0$, $G_1 < -4d_K^*(m_o)m_o s_{K-1} - 2m_o^2 s_{K-1} y_{k-1}/p_K$. Further, G_1 is of the same sign as $G_2 = -2d_K^*(m_o)p_K - m_o y_{K-1} < -(m_o y_{K-1} + d_K^*(m_o)p_K) < 0$. Thus, for $dm_o > 0$ we have $f(d_K^*(m_o) + dd_K, m_o + dm_o) = f(d_K^*(m_o), m_o) = g(d_K^*(m_o), m_o)$ $> g(d_K^*(m_o) + dd_K, m_o + dm_o)$. Thus, if we increase m_o while adjusting d_K to maintain the same level of choice balance (constant a^*), $T' < 0$. Hence, $0 < d_K^*(m_o + dm_o) < d_K^*(m_o) + dd_K$ and $0 < a^*(m_o + dm_o) < a^*(m_o)$. Thus, choice balance increases.

Relative Efficiency of Configurator Questions for E-commerce Decisions

Lemma 3. If e_k, e_ℓ, and e_t are zero-mean, i.i.d. normally distributed , then Prob[$e_k \ge -p_k$] > Prob[$e_\ell \pm e_t \ge -p_k$] for k = 1, 2.

Proof. If the errors are independent and identically distributed zero-mean normal random variables with variances, σ^2, then, for $p_k \geq 0$, the comparison reduces to the following equation where $f(\bullet|\sigma)$ is a normal density with variance σ^2. Let $e_T = e_\ell \pm e_t$.

$$\tfrac{1}{2} + \int_{-p_k}^{0} f(e_k|\sigma)\,de \;\geq \tfrac{1}{2} + \int_{-p_k}^{0} f(e_T|\sigma\sqrt{2})\,de_T \;\Leftrightarrow\; \int_{0}^{c} f(e_k|\sigma)\,de \;\geq \int_{0}^{c} f(e_T|\sigma\sqrt{2})\,de \;\; with\; c \geq 0$$

To demonstrate that the last expression is true, we let $I = \int_0^c f(e_T \mid \sigma\sqrt{2})de =$ $\int \exp(-e^2/4\sigma^2)/(\sigma\sqrt{4\pi})de$. Change the variables in the integration setting $g = e/\sqrt{2}$. Then, $I = \int_0^{c/\sqrt{2}} \exp(-g^2/2\sigma^2)/(\sigma\sqrt{2\pi})dg \leq \int_0^c \exp(-e^2/2\sigma^2)/(\sigma\sqrt{2\pi})de = \int_0^c f(e_T \mid \sigma)de$. The conditions for $p_k \leq 0$ yield equivalent conditions.

Proposition 4. For zero-mean normally distributed (i.i.d.) errors, the warm-up questions (Q1+Q2) provide more accurate estimates of feature choice than do pairs of questions that include a metric paired-comparison question (Q3).

Proof. With Q1+Q2 or Q1+Q3, the error in estimating $\hat{p}_1$ is based directly on e_1 and the probability of correct prediction is Prob[$e_1 \geq -p_1$]. With Q2+Q3 we obtain $\hat{p}_1$ by adding Q2 and Q3 and the error in estimating $\hat{p}_1$ is based on $e_2 + e_t$. The probability of correct prediction is Prob[$e_2 + e_t \geq -p_1$]. Thus, for Feature 1, Q1+Q2 and Q1+Q3 are superior to Q2+Q3 if Prob[$e_1 \geq -p_1$] > Prob[$e_2 + e_t \geq -p_1$]. For Feature 2, similar arguments show that Q1+Q2 and Q2+Q3 are superior to Q1+Q3 if Prob[$e_2 \geq -p_2$] > Prob[$e_1 - e_t \geq -p_2$]. Thus, Q1+Q2 is superior to Q2+Q3 and Q1+Q3 if the condition in the text holds. Finally, by Lemma 3 this condition holds for zero-mean, i.i.d., normal errors.