ROBERT ZEITHAMMER and PETER LENK*

Marketers often analyze multinomial choice from a set of branded products to learn about demand. Given a set of brands to study, the authors analyze three reasons why choices from strict subsets of the brands can contain more statistical information about demand than choices from all the brands in the study: First, making choices from smaller subsets is easier, so it is possible to use more choice tasks when the choice data come from a choice-based conjoint survey. Second, choices from subsets of brands better identify and more accurately estimate the covariance structure of unobserved utility shocks associated with brands. Third, subsets automatically balance the brand shares when some of the brands are less popular than others. The authors demonstrate these three benefits of subsets using a mixture of analytical results and numerical simulations and provide implications for the design of choice-based conjoint analyses. They find that the optimal subset size depends on the model, the number of brands in the study, and the designer's resource constraint. In addition to showing that subsets can be beneficial, this article provides a simulation methodology that helps designers pick the best subset size for their setting.

Keywords: choice models, conjoint, experiment design, choice set, probit

Statistical Benefits of Choices from Subsets

To learn about consumer demand, market researchers often analyze multinomial choice from a set of branded products. For example, the discrete choice conjoint analysis technique analyzes consumer choices from different sets of hypothetical product profiles described by their brands, prices, and other attributes (Louviere, Street, and Burgess 2003). Given a fixed set of brands, a key survey design question arises: How many brands should each choice task include to maximize the demand information contained in the data? From a purely statistical point of view—that is, assuming that standard random utility models accurately portray questionnaire respondents—including all brands in each choice set would seem to provide the most information about demand. In other words, excluding brands from choice sets may seem like throwing away potentially useful data. We analyze statistical properties of standard choice models and find that this intuition is incomplete: Choices from random subsets of the considered brands can be statistically more informative than choices from all the brands.

*Robert Zeithammer is Assistant Professor of Marketing, Anderson School of Management, University of California, Los Angeles (e-mail: robert.zeithammer@anderson.ucla.edu). Peter Lenk is Professor of Operations Management Science and Marketing, Stephen M. Ross School of Business, University of Michigan (e-mail: plenk@umich.edu). Michel Wedel served as associate editor for this article.

There are at least three reasons why a choice-based conjoint survey of branded products can produce more information about demand by offering respondents choices from random subsets of brands than by offering them choices from all brands under study. The most important reason is also the easiest to explain: Choices from smaller sets are easier and faster, so the survey with smaller choice sets can use more choice tasks. Specifically, the decision theory literature (Bettman, Johnson, and Payne 1990) implies that the ease and speed of making a choice are approximately linear in size of the choice set. This result implies that the designer can use more choice tasks as long as the total number of profiles each respondent needs to process remains the same. Under this realistic constraint, we show analytically for the multinomial logit (MNL) model and numerically for the multinomial probit (MNP) model that randomized strict subsets allow sufficiently more tasks to improve demand estimation while keeping the survey time and difficulty perceived by respondents constant.

Coupling smaller subsets with more choice tasks per respondent is not necessary for subsets to be beneficial. We document two reasons why choices from random subsets of the considered brands can be statistically more informative than the same number of choices from all the brands: variance estimation and autobalancing. The variance estimation benefit is specific to models, such as the MNP, that allow for correlated random utilities. Subsets can improve vari-

ISSN: 0022-2437 (print), 1547-7193 (electronic)

ance and covariance estimation because they provide exclusion restrictions (Keane 1992). The variance–covariance terms specific to brands absent from any given task are excluded from the likelihood of that task, allowing that task to better estimate the rest of the parameters. For example, consider three cafés with their respective market shares as follows: Starbucks (50% share), Peet's (25% share), and Bob's (25% share). The Starbucks share advantage could be due to the latter two cafés having weaker brands or to a positive correlation between their random utility shocks. A way to resolve this question is to assume correlations away in the model (e.g., by assuming the MNL) and attribute differences in shares solely to brand partworths. This is the familiar independence of irrelevant alternatives (IIA) assumption (Currim 1982; Huber and Puto 1983; Tversky and Simonson 1993). However, incorrectly assuming IIA has adverse implications for marketing strategy, such as cross-effects of marketing mix or competitor entry (Allenby and Lenk 1994, 1995; McCulloch and Rossi 1994), and so allowing correlations of utility shocks in the model is desirable. Although the aforementioned shares alone do not identify such a model, shares in subsets help disentangle the two possible explanations for Starbucks's share advantage: Suppose that Bob's closes down, and the resultant shares in the subset are 50–50 between Starbucks and Peet's. The correlation of random utilities would explain the full-set shares—perhaps (and unknown to the analyst) Bob's and Peet's are similar local brands. However, if the subset shares are 2:1 in favor of Starbucks, then Peet's café simply has a weak brand (and the shares happen to exhibit IIA). The café example shows that subsets together with the full set improve both identification and estimation. We build on this result and prove that subsets improve identification and estimation even without the data on full-set choice shares. We provide analytical results using simple examples and then conduct a set of simulation experiments to demonstrate the effect on a larger MNP example.

The autobalancing benefit operates whenever the study includes relatively inferior brands that are rarely chosen from the full set of K brands. Subsets can improve estimates of small-brand partworths because inferior brands are chosen more frequently from subsets than from the full set. Kuhfeld, Tobias, and Garratt (1994), Huber and Zwerina (1996), Arora and Huber (2001), and Sandor and Wedel (2001) demonstrate that balancing utilities of the profiles in a choice task can improve estimation precision. They advocate constructing balanced profiles given a fixed choice-set size, which requires prior information about partworths. Balancing by subsets does not require prior information or even qualitative knowledge about which brands are dominated. Instead, useful balancing can occur automatically through randomization, and even completely randomized subsets can be beneficial. To demonstrate the autobalancing benefit, we again show its existence analytically for a small example and then conduct an extensive simulation study to verify generalizability to larger models. Throughout, we document the autobalancing benefit for the MNL that cannot benefit from variance estimation by construction, and thus we show that the two benefits are distinct. We find that using random subsets is optimal whenever more than eight brands are in the study, more than half of them are dominated, few are nonbrand attributes, and brand effects are strong. Unlike the other two benefits, the autobalancing benefit thus only occurs in specific situations. Without the help from one of the other two benefits, autobalancing alone often does not justify the use of subsets.

Our analysis of the three benefits of subsets indicates that smaller subsets are not always more informative than larger subsets. Therefore, pairs are rarely optimal for studies of five or more brands. Instead, the optimal subset size is usually an interior solution, which depends on the particular model, example, objective function, and resource constraint of the analyst. We provide several rules of thumb for choice-based conjoint designers, plus a simulation methodology that can be applied to specific settings. We also conduct a counterfactual simulation of a real-world conjoint analysis by first estimating the parameters of a hierarchical Bayes MNP (HB-MNP) model using data provided by Sawtooth Software and then simulating counterfactual data sets for all possible subset sizes the designer could have used. We find that parameter heterogeneity blunts the benefits of subsets and that the HB-MNP model usually gains less from subsets than the homogeneous MNP. Nevertheless, strict subsets remain optimal under the realistic resource constraint on the total number of profiles per respondents, and double-digit increases in measurement accuracy are possible for both the population-level model parameters and key posterior predictive measures, such as brand equity.

In this article, we assume that conjoint respondents adhere to the assumed random utility models. However, real-world respondents tend to violate these assumptions: People facing large choice sets switch their decision-making strategies from careful compensatory assessment (assumed by standard choice models) to simplifying heuristics (Gilbride and Allenby 2006; Payne, Bettman, and Johnson 1988). Therefore, random utility choice models are more likely to fit actual choice better for smaller subsets, resulting in "cleaner" data (DeShazo and Fermo 2002). These behavioral benefits complement the statistical benefits we discuss here, further strengthening the case for using subsets in choice experiments. Pinnell and Englert (1997) focus on the behavioral benefits and find that pairs are not necessarily better in out-of-sample predictive performance than septuplets, even when the out-of-sample tasks are also pairs. Therefore, the question whether small choice sets are immune from heuristics is by no means settled in conjoint practice.

We motivate our investigation with the example of branded products. Associating random utility shocks with brands is standard in MNP models of consumer choice (Elrod and Keane 1995; McCulloch and Rossi 1994). Our finding generalizes to any categorical attribute with K levels, such that each choice set corresponds one-to-one with a subset of the K levels. For example, transportation studies of commuting behavior often involve K modes of transport such as {train, bus, car, walk}, with each respondent facing a choice from either all the modes or a subset of the modes (Ben-Akiva and Lerman 1985). Note that brand is nevertheless a special attribute because it is difficult to decompose. In the café example, suppose that the designer observes the "localness" attribute that gives rise to the unobserved similarity between Bob's and Peet's. The designer may not be able to control for localness explicitly in the questionnaire (and, thus, to justify i.i.d. errors) because it may be too inherent in the brand identities: The respondents may not be able to imagine a "local Starbucks" or "national leader Bob's."

We organize this article as follows: First, we summarize our theoretical results and develop the reader's intuition for the benefits. Second, we test the variance estimation benefit on a particular homogeneous MNP example. Third, we generalize the model to HB-MNP and use real-world data to calibrate the parameters. We find consistent evidence to support subset benefits due to resource constraints and variance estimation, but the benefit of autobalancing depends on moderating conditions. Therefore, fourth, we set out to better characterize when autobalancing would be expected to operate. The results lead to useful rules of thumb, and they also explain why large benefits from autobalancing are relatively rare. Finally, we discuss the results, their implications for practice, and fruitful directions for future work.

THEORY: TWO CONSTRAINTS AND TWO BENEFITS

Throughout this article, we focus on a specific random utility model structure commonly used in discrete choice analysis (DCA): There are K brands to study and N choice tasks (observations). The nth choice task consists of a subset C_n of the brands $\{1, 2, \ldots, K\}$, and S_n is the cardinality of C_n, where $2 \le S_n \le K$. When brand k is not in C_n, there is no alternative for brand k. For choice task n and brand k in C_n, the random utility U_{nk} is as follows:

$$U_{nk} = \alpha_k + X_{nk}\beta + \varepsilon_{nk}, \tag{1}$$

where α_k is the partworth for brand k, X_{nk} is a row vector of nonbrand attributes of length B, β is a column vector of regression coefficients, and ε_{nk} is a zero-mean random utility term. In this section, we focus on the homogeneous model to derive analytical results about the information content of subsets. In the homogeneous model, the parameters α_k and β are common to all respondents. The MNL assumes that $\{\varepsilon_{nk}\}$ are independent draws from the Gumbel distribution. Common identification constraints set α_K and X_{nK} to zero.

In the MNL, gaining analytical insight into the relationship between subset size and estimation accuracy of α_k and β is possible through the Fisher information matrix. Asymptotic standard errors are the inverse of the Fisher information. To simplify the analysis, we assume that each choice task has the same number of profiles: $S_n = S$, where $1 \le S \le K$. Let I_K^S be the Fisher information matrix associated with random subsets of S brands from a total of K brands. The Fisher information matrix is the sum of task-specific negative Hessians over the N tasks: $I_K^S(\alpha, \beta) = \Sigma_n -H_n(\alpha, \beta \mid C_n)$, where each single-task Hessian depends critically on the following vector of probabilities p_k^S that brand k is chosen from a choice set that contains randomly selected S brands:

$$p_k^S \equiv \Pr\left(k \mid X_n, \alpha, \beta, C_n\right) = \frac{\exp\left(\alpha_k + X_{nk}\beta\right)}{\sum_{j \in C_n} \exp\left(\alpha_j + X_{nj}\beta\right)}. \tag{2}$$

Note that this shorthand notation suppresses the p's dependence on task index n through their dependence on X_{nk}. Given p_k^S, the single-task Hessian is relatively easy to write. It has $(K - 1 - S)$ zero columns (and rows), corresponding to the brands absent from the choice set C_n. Let $\chi_j(C_n)$ be a binary indicator of whether brand j is in C_n; that is, let $\chi_j(C_n) = 1$ iff $j \in Cn$, and $\chi_j(C_n) = 0$ if otherwise. Then, the negative Hessian $-H_n{}^S(\alpha, \beta)$ has the following entries:

$$\begin{aligned}
-H_n^S\left(\alpha_j, \alpha_j\right) &= p_j^S\left(1 - p_j^S\right)\chi_j\left(C_n\right)\\
-H_n^S\left(\alpha_j, \alpha_k\right) &= -p_j^S p_k^S \chi_j\left(C_n\right)\chi_k\left(C_n\right)\\
-H_n^S\left(\alpha_j, \beta'\right) &= p_j^S\left[X_{nj} - E_p\left(X_n\right)\right]\chi_j\left(C_n\right)\\
-H_n^S\left(\beta, \beta'\right) &= \sum_{k \in C_n} p_k^S X'_{nk}\left[X_{nk} - E_p\left(X_n\right)\right],
\end{aligned} \tag{3}$$

where β is a column vector and the last term is a matrix. Furthermore, $E_p(X_n)$ is the p-weighted mean of the X_{nk} vectors present in the task n, so the negative single-task Hessian is just a p-weighted variance matrix of the alternative-specific design vectors present in C_n (for a demonstration, see the Appendix). Specifically, the $(K - 1) \times (K - 1)$ submatrix $-H(\alpha, \alpha)$ is the variance matrix of a $(K - 1)$-dimensional multinomial distribution with probabilities $\{p_k^S\}$.

The determinant (also called "D-criterion") and the trace (also called "A-criterion") of I_K^S are two commonly used scalar summaries from the experimental design literature. In general, analytical evaluations of the determinant and trace of Equation 3 are difficult because of the different X_{nk} vectors entering different tasks. We assume that X_{nk} is zero, and the only parameters to estimate are the $(K - 1)$ brand-specific intercepts α_k. When these are perfectly balanced, $\alpha_k = 0$, and then $p_k^S = 1/S$ for all k. The Fisher information then becomes tractable because the probability does not depend on the particular subset, only on the subset's cardinality S. The following lemma summarizes the information as a function of S:

Lemma 1: Suppose that the MNL involves K brands and no other attributes. When all K brand intercepts have the same value, the determinant and trace of the Fisher information have the following closed form:

$$\det\left[I_K^S\left(\alpha \mid \forall j, k : \alpha_k = \alpha_j, N\right)\right] = N^{K-1}\left(\frac{1}{K}\right)\left[\frac{S-1}{S(K-1)}\right]^{K-1}, \tag{4}$$

and

$$\text{trace}\left[I_K^S\left(\alpha \mid \forall j, k : \alpha_k = \alpha_j, N\right)\right] = N^{K-1}\left(\frac{K-1}{K}\right)\left(\frac{S-1}{S}\right). \tag{5}$$

For a detailed proof, see the Appendix. Equations 4 and 5 lay the groundwork for our analysis of the manager's resource constraint.

The Effect of Resource Constraint: Realistic Versus Conservative

Suppose that the manager running a DCA needs to pay the respondents for the opportunity cost of their time. When the respondents find choices from smaller sets to be easier and faster, smaller subsets can be beneficial because they allow the manager to use more tasks per respondent in the same amount of time. Bettman, Johnson, and Payne (1990) measure the speed of choice in controlled experiments and find that it can be additively decomposed into the time requirements associated with elementary information processes, such as reading, adding, multiplying, and comparing. The compensatory decision process that standard choice models assume is analogous to Bettman, Johnson, and Payne's weighted additive rule, which needs $S(K + B)$ multiplications, $S(K + B - 1)$ additions, and S comparisons to make a choice from S profiles described by up to K brands and B nonbrand attributes. Therefore, the time needed to make a choice from a subset of S profiles increases linearly in S, and the manager faces an implied constraint on the total number of profiles $P = NS$ that each respondent needs to process. Bettman, Johnson, and Payne also show that the time correlates closely with the respondents' subjective assessment of effort. Therefore, a manager concerned with keeping the perceived effort constant will also keep the total number of profiles the same. Bettman, Johnson, and Payne's linearity result clarifies exactly how many more choice tasks are possible with smaller subsets: When a DCA questionnaire is limited to P total profiles across all questions, there can be $N = P/S$ tasks with S profiles each.

In contrast, if making a choice from fewer profiles is not easier for the respondents than a choice from more profiles, the designer of the study faces a constraint on the number of tasks (choice observations) per respondent. This constraint is more conservative than the realistic constraint grounded in data on consumer behavior. The simple intercept-only case of the MNL discussed in Lemma 1 can be used to demonstrate that these two resource constraints can have different implications for the usefulness of subsets.

> P_1: Under the assumptions of Lemma 1 and for both the trace and the determinant criterion, the manager constrained to a constant number of choice tasks per respondent learns the most about demand from K-tuples as choice sets. In contrast, the manager constrained to a constant number of profiles learns the most from pairs.

This proposition follows from Lemma 1. When N is held fixed, both the determinant and the trace in Equations 4 and 5 are functions of $(S - 1)/S = 1 - 1/S$, which increases in S. When P is held fixed and there are $N = P/S$ choice tasks, the situation reverses: The determinant becomes proportional to $[(S - 1)/S^2]^{K-1}$ and the trace to $(S - 1)/S^K$, both of which are decreasing in S.

P_1 illustrates that subsets are more likely to be beneficial under the realistic constraint on the number of profiles than under the conservative constraint on the number of tasks. Notably, we find that subsets can be beneficial even under the conservative constraint. This finding is somewhat counterintuitive because using only S brands per task is the same as reducing the number of observations in the latent regression of U on X and brand dummies. If the manager facing a constraint on number of tasks could observe the utilities U_{nk} directly, he or she would never prefer a strict subset size $S < K$. As a result of the peculiar properties of choice data, subsets can benefit estimation even under the stronger constraint of constant N. Subsets can be beneficial in this case for two reasons: variance estimation and autobalancing. We discuss the two reasons in turn.

Variance Estimation Benefit in MNP

Unlike the MNL, which assumes away correlations of the utility shocks $\varepsilon_{n,k}$, MNP assumes that the shocks are draws from a multivariate normal distribution: $(\varepsilon_{n,1}, \ldots, \varepsilon_{n,K-1}) \sim N(0, \Sigma)$. For identification, MNP fixes $U_K = 0$ and $\sigma_{11} = 1$ (McCulloch and Rossi 1994). Given this identification strategy, allowing the Kth alternative to be the outside alternative is natural. The $(K - 1) \times (K - 1)$ covariance matrix Σ is usually interpreted as a measure of unobserved similarities between pairs of "inside" brands.

Strict subsets can be beneficial to the estimation of Σ even under the more conservative constraint (constant number of choice tasks) because they imply exclusion restrictions in the likelihood. A simple trinomial ($K = 3$) example illustrates this point by showing that strict subsets improve identification. Consider the MNP with two brands (i.e., "inside alternatives"), A and B, plus one outside alternative C. Let α_A and α_B be the partworths for A and B. Assume that the error variances are $\sigma_{AA} = \sigma_{BB} = 1$ and that the error covariance is $\sigma_{AB} = \rho$. Assume further that no nonbrand attributes exist. Therefore, we need to estimate only three parameters: $\{\alpha_A, \alpha_B, \rho\}$. In this example, a striking benefit of subsets arises:

> Example 1: The trinomial probit model with two brands, without nonbrand attributes, and with $\sigma_{AA} = \sigma_{BB} = 1$ is not identified when the choice sets are triples ($S = K = 3$), but it is identified when the choice sets are random pairs ($S = 2$) and the two brands are not identical ($\alpha_A \neq \alpha_B$).

The algebraic proof is available in the Appendix, but its key ideas can be easily cast in the café story, where A = Peet's, B = Bob's, and C = Starbucks. When all three cafés are in the choice set, the model cannot be identified, because choices from the full {A, B, C} choice set reveal only two nonredundant pieces of information: share of A and share of B. These two pieces of information cannot separately identify the three parameters of the model. Now consider the choices from the three possible pairs of cafés, each providing a separate piece of information for the estimation: The share of A in {A, C} is increasing in α_A, and it does not depend on any other parameters. Therefore, the share of Peet's café in a market without Bob's café pins down the value of Peet's brand uniquely. Analogously, the share of Bob's café in a market without Peet's café pins down the value of Bob's brand. Finally, only the shares in {A, B} depend on the correlation ρ. When $\alpha_A = \alpha_B$, the share of A in {A, B} is 1/2 for every ρ, so ρ is not identified. However, when $\alpha_A \neq \alpha_B$ and one brand is even slightly preferred (WLOG $\alpha_A > \alpha_B$), it can be shown that its share increases with ρ. Intuitively, as ρ increases, the errors $\varepsilon_{n,A}$ and $\varepsilon_{n,B}$ contain a greater common component, which cancels out the utility comparison that determines the choice: $U_{n,A} > U_{n,B} \Leftrightarrow \alpha_A - \alpha_B > \varepsilon_{n,B} - \varepsilon_{n,A}$.

If the data consisted of repeated observations of shares for {A, B, C} along with sufficiently varying nonbrand attributes $\{X_{nk}\}$, the designer could identify the model for $\{\alpha_A, \alpha_B, \beta, \Sigma\}$ even by full subsets. Though less stark than in Example 1, the improved and more direct identification benefit of subsets would remain even in such a standard model with nonbrand attributes. A way to interpret the benefit is that subsets make practical the semiparametric "identification at infinity" argument of Chamberlain (1986).

The foregoing argument for improved identification with subsets generalizes beyond the case of pairs in a trinomial probit. The argument actually simplifies as a result of a combinatorial explosion: A K-nomial intercept-only probit with unit diagonal of Σ has K choose S subsets of size S but only K choose 2 parameters: K – 1 intercepts and K – 1 choose 2 correlations. Therefore, the number of identifying equations rises faster than the number of parameters as K increases. In particular, the unit diagonal of Σ is not necessarily required for identification when $K \geq 6$ and $S = 3$.

This subsection has analyzed the benefit of subsets for identification of MNP parameters. We expect that improved and more direct identification should lead to more precise estimation of the model parameters, but analytical results on estimation precision are difficult to obtain for the MNP. Therefore, we turn to simulations to check the intuition and gauge the magnitude of the benefit, while also considering $K > 3$ and $X_{nk} \neq 0$. Before describing our simulation results, we propose a second benefit of subsets unrelated to variance estimation.

Autobalancing Benefit

In this subsection, we show that subsets can still benefit the estimation of standard choice models even when the covariance structure of ε_{nk} is assumed (as in the MNL) rather than estimated. The subsets then benefit by automatically evening out the overall choice shares across tasks, so we call the improvement an autobalancing benefit. We develop the argument using the MNL because we can express its estimation accuracy in closed form, but we expect the benefit to also hold in an MNP with a fixed and known Σ.

In the binary logit (K = 2), the Fisher information matrix is as follows:

$$I_{Logit}(\alpha,\beta) = \sum_{n=1}^{N} p_n(1-p_n)\begin{bmatrix} 1 & X_n \\ X_n' & X_n'X_n \end{bmatrix},$$

$$\text{where } p_n = \frac{\exp(\alpha + X_n\beta)}{1+\exp(\alpha + X_n\beta)}.$$

Here, p_n is the probability of Option 1 given X_n—a row vector of nonbrand attributes. If we actually observed the latent utilities $U_n = \alpha + X_n\beta + \varepsilon_n$, where $\varepsilon_n \sim N(0, 1)$, the Fisher information would be

$$I_{Reg}(\alpha,\beta) = \sum_{n=1}^{N}\begin{bmatrix} 1 & X_n \\ X_n' & X_n'X_n \end{bmatrix}.$$

A comparison of the two information matrices implies that at least three-quarters of the information is lost in the choice data relative to the latent linear regression, and how much is lost depends on how close p_n is to 1/2, where $p_n(1 - p_n)$ attains its maximum. How does this intuition generalize to multiple-choice models?

In MNL, the Fisher information takes an appealing form with a slight change of notation. Let $p_{k,n}$ be the probability of choosing brand k in choice task n, and let $\varphi = (\alpha_1, \ldots, \alpha_{K-1}, \beta')'$ be the full coefficient vector (so the utility of k is $w_{k,n}\varphi$, where $w_{nj} = [0, \ldots, 0, 1, 0, \ldots, 0, x_{nj}]$ is a K – 1 + B row vector of both the brand and nonbrand attributes). The MNL Fisher information becomes

$$(6) \qquad I_{MNL}(\varphi) = \sum_{n=1}^{N} W_n'(P_n - \vec{p}_n\vec{p}_n')W_n,$$

$$\text{where } W_n = \begin{bmatrix} w_{n1} \\ \vdots \\ w_{nK} \end{bmatrix}, P_n = \text{diag}(p_n), \text{and } \vec{p}_n = \begin{bmatrix} p_{n1} \\ \vdots \\ p_{nK} \end{bmatrix}.$$

If we actually observed the latent utilities $U_n = W_n\varphi + \varepsilon_n$, where $\varepsilon_n \sim N_K(0, I)$, the Fisher information would be the information matrix of a multivariate regression:

$$(7) \qquad I_{MVR}(\varphi) = \sum_{n=1}^{N} W_n'W_n.$$

As in the case of the binary logit, the comparison of I_{MVR} with I_{MNL} reveals the information loss due to only observing choice: The kernel $P_n - \vec{p}_n\vec{p}_r'$ in I_{MNL} tends to become small (trace = $1 - p_1^2 - \ldots - p_K^2$) compared with the identity matrix, resulting in a large degradation in information for larger subset sizes. Can smaller choice sets actually provide more information with the same number of tasks? To answer these questions, we first consider the simpler intercept-only case. The next example demonstrates that the intercept-only MNL can benefit from subsets, even under the conservative resource constraint when different brands have different values.

Example 2: Suppose that the MNL involves K = 3 brands and no other attributes. Fix $\alpha_3 = 0$ for identification. Based on the determinant of the information matrix, the model is more precisely estimated with N random pairs than with N triples whenever Brands 1 and 2 are sufficiently dominated, such that $\alpha_1 \leq \alpha_2 < -(5/2)$. Based on the trace of the information matrix, the model is more precisely estimated with N random pairs than with N triples for all α_1 and α_2.

The example is worked out in the Appendix. Recall that in P_1, the model is estimated more precisely with N full choice sets of size K than with N random subsets of size S given that the intercepts are equal to zero. Example 2 is different from the case in Lemma 1 and P_1 in that two of the brands are dominated by the outside good or Brand 3. When $\alpha_1 = \alpha_2 = -(5/2)$, the share for Brand 3 is 86%. The key intuition for why pairs are better than triples is that when the choice task consists of the dominated brands, the probabilities are close to 1/2, which maximizes the Fisher information in that task. When the choice set is all three brands, the dominated brands contribute little to information. As the information is added over tasks, the pairs that happen to exclude the dominant alternative can more than compensate for the loss of information arising from the pairs that include it (and thus are even less informative than the full triple). This benefit is available even to a manager who does not know which brand is the dominant one—thus, the "auto" in our name for this benefit. If the manager knows that Brand 3 is the dominant one, nonrandom subsets will be beneficial more often. For example, it can be shown that based on the determinant criterion, excluding the dominant Brand 3 from half the tasks will be beneficial whenever $\alpha_1 \leq \alpha_2 < \log(1/4) = -1.38 > -(5/2)$.

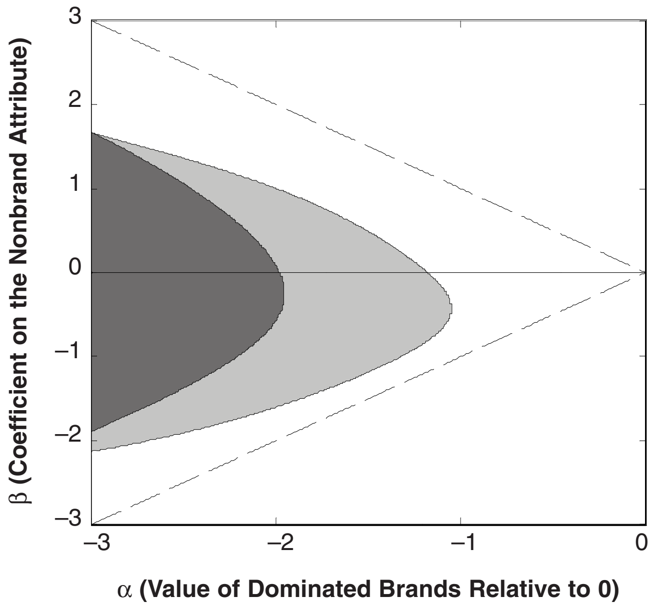
The simple intercept-only model is unrealistic, but it enables us to make rigorous statements about the estimation properties of the MNL. Two questions arise: First, how is the "S = K is optimal when brands are equal" result related to the utility-balancing idea? Second, how does the inclusion of nonbrand attributes in the model influence the optimality of subsets? The first question can be answered as follows: Fix S = K, and allow brand values α_k to vary. The utility-balancing literature would ask what the distribution of brand values is that maximizes information when S = K. The answer would be that equal brand values are optimal because $p_j = (1/K)$ for j = 1, …, K maximizes $\det(I_K) = N^{K-1}\Pi_{j=1}^{K}P_j$. This focus on the selection of brands given a choice set size is different from our focus on the selection of the best choice set size given brands. The second question is more difficult to tackle analytically, but showing that subsets can be more informative even with nonbrand attributes is possible, especially when dominated brands exist. To show this claim, we construct a third example.

Example 3: Suppose the MNL involves three brands and one binary nonbrand attribute X_n, which is randomly equal to 1 for exactly one of the alternatives in C_n and equal to 0 for the other alternatives in C_n. Let $\alpha_3 = 0$ and $\alpha \equiv \alpha_1 = \alpha_2$. Then, there is a region of the (α, β) parameter space in which N random pairs are more informative than N triples (see Figure 1).

To construct the information matrix in Example 3, all combinations of brands and nonbrand attributes must be considered, which is exponentially difficult, thus limiting analytical tractability. Figure 1 shows that the conclusion of Example 2 is nevertheless robust to the inclusion of nonbrand attributes: When Brands 1 and 2 are sufficiently dominated, pairs prevail over triples. As in Example 2, the trace criterion makes subsets optimal more often, but not always. Note that the case of $\beta = 0$ is not identical to Example 1, because β is still estimated here instead of being fixed to 0 as in the intercept-only case: For example, the boundary of the darker-shaded (determinant criterion) area intersects the α axis around –2, whereas Example 1 requires $\alpha < -(5/2)$ for pairs to be optimal. In this sense, nonbrand attributes do not necessarily weaken the autobalancing benefit of subsets when their effect is weak. However, Figure 1 clearly shows that nonbrand attributes with larger partworths weaken the autobalancing benefit substantially: More dominated brands are needed for pairs to be preferred to triples as $|\beta|$ rises.

The simple examples of this section show that subsets are preferred to complete choice tasks when (1) the total number of profiles is held constant in the study, (2) subsets provide exclusion restrictions that improve the estimation of variances and covariances, and (3) subsets provide autobalancing of utilities when there are dominant brands. In the remainder of the article, we use numerical simulations to assess their likely incidence and magnitude in practice.

Figure 1
WHEN ARE RANDOM PAIRS MORE INFORMATIVE THAN TRIPLES IN A TRINOMIAL LOGIT?



Notes: The model has 3 brands and 1 nonbrand binary attribute; the value of the dominant brand is fixed at zero. The shaded areas show the regions in which random pairs are more informative than full triples: The darker area is based on the determinant criterion, and the lighter area is based on the trace criterion.

VARIANCE ESTIMATION BENEFIT: SIMULATION RESULTS FOR THE MNP

Method

The "Theory" section shows that choices from random subsets can identify the parameters of an MNP better and more directly than choices from full sets. We expect the improved identification to accompany improved accuracy of the parameter estimates. This section investigates our hypothesis using simulations because closed-form results about MNP estimation errors are not available. The simulations follow a 2 × 2 experimental design that varies which constraint the manager is facing (fixed number of profiles versus fixed number of choice tasks) and whether Σ is fixed or estimated. We use the fixed Σ case to mimic the MNL.

The model example we simulate consists of five brands and one outside option (K = 6). The brand intercepts are $\alpha_j = -.5$ for j = 1, …, 5 and $\alpha_6 = 0$. We assume that the random components of utility are multivariate normal with the covariance matrix Σ set to have varied diagonal elements between .6 and 1.2 and varied correlations between –.5 and .5. In addition to the brand intercepts, one randomly generated binary attribute X_{nk} is meant to represent an indicator of brands being on sale $[\Pr(X_{nk} = 1) = 1/2]$. Its regression parameter is $\beta = 1$. We set $U_6 = 0$ and $\sigma_{11} = 1$ for model identification.

For every subset size S = 2 to S = 6, we generate synthetic tasks from the model and estimate the parameters using the Gibbs sampler proposed by Zeithammer and

Lenk (2006). We set the conservative constraint to N = 2000 tasks and let the realistic constraint on the number of profiles P = NS coincide with the conservative constraint in the full S = 6 case. Therefore, the number of tasks under the weak constraint increases proportionally from N = 2000 when S = 6 to N = 6000 when S = 2. To measure the information content of different subset sizes, we calculate the following posterior accuracy measures:

- Standard deviation of the posterior distribution of each scalar parameter $\theta = \{\alpha_k, \beta, \sigma_{jk}\}$.
- Mean square error (MSE) between true and predicted choice probabilities in a full-set (S = K = 6) holdout sample. We compute the MSE in a single holdout task as follows:

$$\text{MSE} = \frac{1}{KM}\sum_{k=1}^{K}\sum_{m=1}^{M}\left[\Pr_k\left(\hat{\theta}_m\right) - \Pr_k\left(\theta_{true}\right)\right]^2,$$

where $\hat{\theta}_m$ is the mth draw of M total Markov chain Monte Carlo iterations from the posterior of θ.
- Standard deviation of two important posterior predictive distributions:

1. "Brand equity" is the change in share associated with setting α_1 to zero. In a single task,

$$\text{Brand equity} = \frac{1}{M}\sum_m\left[\Pr_1\left(\hat{\theta}_m \text{ with } \alpha_1 = 0\right) - \Pr_1\left(\hat{\theta}_m\right)\right].$$

2. "Sales lift" is the change in share associated with increasing the utility of the first alternative by β. In a single task,

$$\text{Sale lift} = \frac{1}{M}\sum_m\left[\Pr_1\left(\hat{\theta}_m \mid X_{n1} = X_{n1} + 1\right) - \Pr_1\left(\hat{\theta}_m \mid X_{n1}\right)\right].$$

For both the MSE and the posterior predictive measures, we average over the holdout tasks. The first set of measures enables us to study the effect of subsets on estimation accuracy in utility space, the second set of measures combines the parameters together and translates them into the probability space but considers only the estimation of levels, and the third set of measures evaluates the changes in probabilities (i.e., changes in shares) as a function of changes in the independent variables.

Counterfactual questions, such as "How would our share change if we did not have our brand?" or "How much would our share rise if we promoted our product more?" are often the reason companies conduct conjoint analyses (Orme 2005). Therefore, we believe that the differences-in-probabilities measures are the most managerially relevant because they represent answers to counterfactual questions that are difficult to assess without a model. For example, brand equity—"How would our share change if we did not have our brand and everything else remained the same?"—is one of the most important counterfactual questions that marketers ask. Our measure of brand equity answers it directly by focusing on market share. A direct antecedent of brand equity in the utility space is α_1—Kamakura and Russell's (1993) "brand value" measure. Our measure is different from that of Goldfarb, Lu, and Moorthy (2009), who ask the even deeper question of "How much of our share is due to our brand?" To answer that question, they need to model a change in equilibrium prices after setting α_1 to zero, while we keep all X variables constant in the counterfactual simulation.

Given a design X and a set of parameters θ, we need to integrate over the particular data and subset realizations to measure the expected estimation accuracy and assess the random variation around this average. To achieve this integration, we repeat the simulation 100 times for every subset size. The exact algorithm we use is as follows:

Initialization: Fix values of the parameters θ, priors on all parameters (we use standard diffuse priors from the literature), and the managerial constraint (number of tasks or number of profiles). Then, repeat the following 100 times (computational details are in the Web Appendix at http://www.marketingpower.com/jmrdec09):

1. Draw N full K-dimensional utilities U_n and a design X_n for calibration and prediction.
2. For each subset size, S = 2, …, K.
   a. Randomly select S of the K "brands" for each task n. The outside alternative K is treated the same as any other alternative, so some subsets do not include it.
   b. Simulate the choices according to maximum utility for each choice set.
   c. Use Markov chain Monte Carlo to generate draws from the posterior distribution of θ.
   d. Estimate θ and compute predicted probabilities and performance measures.

Results

Table 1 shows the results of our simulation study. Each row of the table is normalized such that "100%" corresponds to the full data case of S = K = 6 and N = 2000, which corresponds to P = 12,000 profiles. The numbers are average normalized accuracy measures; that is, we first divide measure S by measure K within each repetition and then average across repetitions. For completeness, Table W1 in the Web Appendix (http://www.marketingpower.com/jmrdec09) reports the actual average values of the normalizing cases (S = K). In general, the normalizing values are small because we selected moderately large designs that estimate the models quite well.

A few general impressions emerge from Table 1. Except for the (Σ known, constant number of tasks N) case, subsets provide benefits, and the optimal numbers of subsets are usually interior solutions—not pairs or complete sets. The benefits are greatest in the (Σ estimated, constant number of profiles) case, in which all three benefits of subsets operate together. The optimal subsets depend on the accuracy measure, with triples "winning" the most often. The (Σ estimated, constant number of tasks) case illustrates the variance estimation benefit explored in the "Theory" section. As we predicted, brand intercepts and covariance parameters benefit robustly from subsets, with precision improvements of 15%–30%. As we also expected, a clear trade-off exists between the estimation of brand intercepts α and attribute partworths β: The parameter on nonbrand attributes benefits from subsets much less, and relatively larger subsets estimate the parameter best. The (Σ fixed, constant number of tasks) case reveals that the autobalancing benefit does not operate in the region of the parameter

Table 1

VARIANCE ESTIMATION BENEFIT IN MNP: THE EFFECT OF MANAGERIAL RESOURCE CONSTRAINT

	Σ Estimated, Constant Number of Tasks				*Σ Estimated, Constant Number of Profiles*			
	Subset Size (Number of Tasks)				*Subset Size (Number of Tasks)*			
Average Standard Deviation of Posterior (Subsets = 100%)	*2 (2000)*	*3 (2000)*	*4 (2000)*	*5 (2000)*	*2 (6000)*	*3 (4000)*	*4 (3000)*	*5 (2400)*
Deterministic U Parameters α, β								
α (brand intercepts)	**80%**	***70%***	**75%**	**87%**	**50%**	***49%***	**63%**	**80%**
β (binary attribute parameter)	142%	**97%**	***86%***	**88%**	**91%**	***71%***	**73%**	**82%**
Random Utility Parameters Σ								
Variances	187%	103%	**90%**	***89%***	105%	***74%***	**77%**	**84%**
Correlations	164%	**91%**	***77%***	**80%**	**111%**	**67%**	***66%***	**74%**
Predicted Difference in Share								
Sale lift of Brand 1	141%	108%	100%	***99%***	**99%**	***80%***	**84%**	**99%**
Brand equity 1	**74%**	***61%***	**70%**	**88%**	***42%***	**44%**	**59%**	**79%**
Overall Holdout Performance (MSE, No Subsets = 100%)								
MSE (true versus predicted probability)	689%	188%	115%	***94%***	281%	**93%**	***78%***	**82%**
	Σ Known, Constant Number of Tasks				*Σ Known, Constant Number of Profiles*			
Deterministic U Parameters α, β								
α (brand intercepts)	160%	122%	108%	101%	**92%**	***85%***	**87%**	**93%**
β (binary attribute parameter)	158%	123%	110%	103%	**90%**	***86%***	**89%**	**94%**
Predicted Difference in Share								
Sale lift of Brand 1	168%	126%	112%	105%	104%	***93%***	**98%**	**103%**
Brand equity 1	161%	121%	109%	102%	**86%**	***84%***	**88%**	**93%**
Overall Holdout Performance (MSE, No Subsets = 100%)								
MSE (true versus predicted probability)	298%	186%	138%	109%	110%	100%	**99%**	***96%***

Notes: Except for the overall holdout performance, the table reports standard deviations of posterior distributions normalized such that the "full set" (S = 6 and N = 2000) is 100%. We first normalize for every repetition and then compute the averages shown. The standard errors of the means are all less than 1% for parameters, less than 3% for the posterior predictive probabilities, and rising from 2% to more than 20% for the normalized MSE as the subset size decreases from 5 to 2. Bold font represents benefits from subsets, and the italicized bold numbers show the best subset size for every statistic of interest, whenever there is a benefit from subsets.

space we consider here. In the next section, we corroborate this explanation by finding that at least in the MNL, strong-enough autobalancing should not be expected with K = 6 and small α. Finally, the two (Σ fixed) cases isolate the main effect of the managerial resource constraint. The finding echoes the MNL result that P_1 discusses: Compared with keeping the number of tasks constant, smaller subsets are optimal when the manager needs to keep the number of profiles constant. For such a manager, the optimal subsets are not pairs as the stylized MNL example would predict, because the model includes a nonbrand attribute. Notably, the optimal subset size depends on the measure of posterior precision in which the manager is interested, with smaller subsets (triples) optimal and relatively more beneficial for the precision of parameters and probability changes and larger subsets (quintuples) optimal and relatively less beneficial for the precision of the level of probability.

A consistent result is that probability changes (sales lift and brand equity) tend to benefit more from subsets than probability levels (holdout predictions). An explanation for this pattern can be gleaned from the trinomial café example we discussed previously: The shares from the full market consisting of {A, B, C} are good estimates of the levels of shares in the same market tomorrow. However, making a prediction about what would happen in a counterfactual market, in which A lost the brand, is difficult. In the trinomial example discussed in the "Theory" section, this prediction is actually impossible to identify. Given this intuition and the simulation results, we conclude that subsets tend to be more helpful in predicting counterfactual probability changes than in predicting probability levels. The comparison of "sale lift" with "brand equity" clarifies that the predictions of probability changes involving changing brand utility values benefit more than predictions involving only changing nonbrand attributes. The key role of Σ in the former counterfactual drives this difference because the difference is present only when Σ is estimated (note that in the fixed-Σ cases, the precision of the two probability changes is a function of subset size).

The orthogonal design enables us to assess the interaction between the benefit from variance estimation and the benefit from the looser realistic constraint. We focus on the posterior predictive measures (brand equity, sale lift, and holdout fit), and we find almost no interaction in the sense that for every subset size S, the precision improvement due to the realistic constraint is approximately the same when Σ is estimated as when it is fixed. An exception to this pattern is the brand equity measure, which benefits from the realistic constraint less when Σ is estimated.

APPLICATION: HB-MNP DCA OF DEMAND FOR PERSONAL COMPUTERS

Method

In this section, we test the extent to which the benefit of subsets identified for homogeneous MNP models carry over to HB-MNP with realistic heterogeneity. To achieve this goal, we fit a HB-MNP model to a real DCA study and simulate counterfactual data sets with different subset sizes as if the estimated parameters were the true parameters. We use the same 2×2 experimental design as in the previous section. Sawtooth Software provided the data for the conjoint study, which involved 316 information technology purchasing managers choosing personal computers. The purchasing managers had five brands from which to choose, plus the outside alternative (K = 6), and each respondent was presented with eight choice tasks. Each choice task consisted of three computer profiles, with each brand appearing, at most, once in each choice task. Therefore, with the "none-of-the above" outside option included in every task, the choice sets were quadruples (S = 4). In addition to the five brands, the study manipulated four different levels of price and three different levels of performance, warranty, and service attributes. In general, brands A and B are considered premium brands, and brands D and E are value brands.

The key difference between the homogeneous models of the previous section and the HB-MNP is respondent heterogeneity in parameters captured by a hierarchical prior (Lenk et al. 1996). Specifically, if the jth respondent has utility coefficients $[\alpha_j, \beta_j]$, we let $[\alpha_j, \beta_j] \sim N([\alpha, \beta], \Lambda)$. We assume that Λ is a diagonal matrix because with our sample size, estimates of the full covariance matrix are sensitive to the prior assumptions. The several levels of nonbrand attributes are coded as continuous variables normalized between 0 and 1, so we measure their coefficients in the units of utility (i.e., the standard deviation of the first-brand random component). The population-level estimates are in Table 2. The brand intercepts are ordered as expected but are not dramatically different from one another. The estimate of Σ reveals that brands {A, B, C} are relatively similar to one another, and brand E is relatively dissimilar from them. This finding confirms our intuitive understanding of the similarities between the brands. The respondents are (on average) price sensitive, and the most important attribute is performance. We find a lot of heterogeneity in price sensitivity and the first-brand intercept.

We modify the simulation procedure in the last section to conduct the counterfactual simulation analysis with modifications for parameter heterogeneity. We generate 316 individual-level parameters $[\alpha_j, \beta_j]$ from the estimated distribution of heterogeneity. The complete choice task size is K = 6 with five brands and the outside good. For each respondent, we generate P = 60 profiles. From these 60 profiles, we randomly construct choice tasks of various sizes S = 2, …, 6, such that each brand appears, at most, once in each choice task. Under the constant N constraint, ten choice tasks exist, and $P = 10 \times S$. Under the constant P constraint, 60 profiles exist, and N = P/S. We also generate one holdout task per respondent for prediction. Next, we use the individual parameters $[\alpha_j, \beta_j]$ and the estimated error variance matrix Σ to generate new utilities $U_{j,k|t}$ and the discrete choice data. Then, we estimate the HB-MNP model with both estimated and known Σ and calculate the posterior measures. We calculate the holdout MSE and changes in probabilities at the individual level and then average across respondents. Therefore, the changes in probabilities correspond to predicted changes in share while taking heterogeneity into account. We repeated this exercise 100 times to obtain the numbers in Table 3 in the same way as Table 1.

Results

Table 3 presents the results in the same format as Table 1. Note that Table 3 focuses on the population-level parameters $[\alpha, \beta]$, not on the individual-level parameters $[\alpha_j, \beta_j]$. The Bayesian estimation procedure for HB-MNP enables us to examine the posterior distributions of the individual-level parameters as well. Specifically, we let the first person actually have $[\alpha_1, \beta_1]$ equal $[\alpha, \beta]$ throughout the simulations. We find that the pattern of gains is in the same direction for the individual-level $[\alpha_1, \beta_1]$ and the population-level parameters $[\alpha, \beta]$, but the benefits and costs are muted for the first respondent (contact the authors for details). A deeper analysis of individual-level estimation is beyond the scope of this article, but note that we compute all posterior predictive probabilities at the individual level, so the brand equity, sales lift, and holdout measures fundamentally stem from the individual-level estimates.

Table 3 confirms the pattern of benefits found in Table 1, but the benefits are not as prominent for HB-MNP as for the homogeneous MNP. Under the conservative constraint (constant number of tasks), we find that subsets benefit only the estimation of Σ parameters and the precision of the "brand equity" probability difference. The optimal subsets for both are relatively large (quintuples). The variance esti-

Table 2

PARAMETER ESTIMATES IN THE PERSONAL COMPUTER STUDY: MEAN OF THE POSTERIOR DISTRIBUTION

	Deterministic Utility Parameters		*Variances and Correlations of the Random Utilities*				
	Population Mean	*Population Standard Deviation*	*brandA* ε	*brandB* ε	*brandC* ε	*brandD* ε	*brandE* ε
brandA α	–1.74	.86	1.00				
brandB α	–2.03	.50	.24	1.09			
brandC α	–2.27	.46	.15	.21	1.23		
brandD α	–2.54	.46	.15	.03	–.02	1.15	
brandE α	–2.87	.69	–.18	–.17	.29	.05	.98
price β	–1.58	.82					
performance β	2.70	.77					
warranty β	1.14	.52					
service β	1.27	.53					

Notes: The nonbrand attributes are scaled to [0, 1].

Table 3

COUNTERFACTUAL BENEFIT OF SUBSETS IN THE PERSONAL COMPUTER STUDY (HB-MNP MODEL)

	Σ Estimated, Constant Number of Tasks				*Σ Estimated, Constant Number of Profiles*			
	Subset Size (Number of Tasks/Respondent)				*Subset Size (Number of Tasks/Respondent)*			
Standard Deviation of Posterior, (No Subsets = 100%)	2 (10)	3 (10)	4 (10)	5 (10)	2 (30)	3 (20)	4 (15)	5 (12)
Deterministic Utility Parameters α, β								
Population Mean								
α (brand intercepts)	178%	128%	108%	101%	105%	**89%**	***88%***	**92%**
β (attribute parameters)	198%	140%	115%	104%	118%	**98%**	***95%***	***95%***
Deterministic Utility Parameters α, β								
Population Standard Deviation								
α and β	197%	141%	118%	106%	110%	***95%***	***95%***	**96%**
Random Utility Parameters Σ								
Variances	178%	116%	**92%**	***83%***	**96%**	**78%**	***75%***	**76%**
Correlations	145%	108%	**94%**	***88%***	**97%**	**79%**	***78%***	**81%**
Predicted Difference in Share								
Sale lift of Brand 1	164%	125%	112%	105%	102%	***94%***	**95%**	**98%**
Brand equity 1	167%	119%	104%	***99%***	**97%**	***85%***	***85%***	**93%**
Overall Holdout Performance								
(MSE, no subsets = 100%)								
MSE (true versus predicted individual probability)	165%	136%	118%	108%	113%	100%	**98%**	99%
	Σ Known, Constant Number of Tasks				*Σ Known, Constant Number of Profiles*			
Deterministic Utility Parameters α, β								
Population Mean								
α (brand intercepts)	154%	124%	111%	104%	**93%**	***89%***	**91%**	**96%**
β (attribute parameters)	152%	124%	112%	105%	**96%**	***92%***	**94%**	**97%**
Deterministic Utility Parameters α, β								
Population Standard Deviation								
α and β	152%	127%	114%	106%	**95%**	***90%***	**93%**	**97%**
Predicted Differences in Share								
Sale lift of Brand 1	163%	128%	113%	106%	102%	***95%***	**96%**	**99%**
Brand equity 1	170%	128%	112%	104%	**98%**	***91%***	***91%***	**97%**
Overall Holdout Performance								
(MSE, no subsets = 100%)								
MSE (true versus predicted individual probability)	157%	133%	117%	107%	113%	100%	***98%***	**99%**

Notes: Except for the MSE, the table reports standard deviations of posterior distributions normalized such that the "full set" (S = 6) is 100%. We first normalize for every repetition and then compute the averages shown. The standard errors of the means are all less than 2% except for MSE in the case of very low S. Bold font represents benefits from subsets, and the italicized bold numbers show the best subset size for every statistic of interest, whenever there is a benefit from subsets. N = 316. The normalization in each row is the same for both the constant tasks and constant profiles case: 100% = statistic with 10 choice tasks per respondent, with S = 6.

mation benefit still operates in that subsets are never optimal in the known Σ case, whereas they are optimal at least for some measures when Σ is estimated. Though less compelling to a manager facing a conservative constraint, the benefits of subsets are still present in that the accuracy does not decline quickly as S decreases. For example, the sale lift measure is only 5% less precise with quintuples than with full sets.

Because subsets improve the estimation of Σ more than other parameters, we consider an additional managerially relevant counterfactual issue that is especially sensitive to Σ—namely, a competitor's exit. Specifically, recalling the three cafés example, correctly predicting the increase in A's share due to B's exit depends on correctly estimating the correlation of their random utilities. The results in Table 3 suggest a comparison between quintuples (optimal for Σ) and sextuples (optimal for other parameters). We compute the appropriate comparisons and report only the overall findings in this paragraph. We make two comparisons: individual and aggregate. On the individual level, the question is whether the holdout MSE between true and predicted probabilities (of choosing A without B in the market) is lower with quintuples or with full sets. We find that this is not the case; the MSE is 10% higher with quintuples. Therefore, even a double-digit improvement in the estimation accuracy of Σ does not compensate for a single-digit reduction in precision of α when it comes to predicting individual-level probabilities. However, we find that the posterior standard deviation of the change in share of A as a function of B's exit mirrors closely the behavior of the brand equity measure. Therefore, the large improvement in the estimation of Σ slightly improves the accuracy of the aggregate change-in-share measure. We suspect that the difference between the benefit of subsets of $[\alpha_j, \beta_j]$ and $[\alpha, \beta]$ and the nonlinearity of the probability function drive the individual versus aggregate difference.

The effect of weakening the manager's resource constraint remains almost exactly as strong in Table 3 as in Table 1. Therefore, a manager who faces the realistic constraint benefits robustly from subsets even when substantial heterogeneity exists. Echoing the results in Table 1, pairs are not the optimal subsets for any measure under consideration, strengthening the evidence against their use in practice. Again, the interaction between variance estimation and constraint weakening is small: The manager who faces the realistic constraint is better off than the manager who faces the stronger conservative constraint, whether they both estimate Σ or assume that it is known, as in MNL.

AUTOBALANCING BENEFIT: SIMULATION RESULTS FOR THE MNL

Method

The trinomial logit Examples 2 and 3 in the "Theory" section prove that subsets can be beneficial in MNL estimation due to autobalancing, even under the conservative resource constraint. In the two sections for the MNP and the HB-MNP, the test of autobalancing is the known Σ cases because variance estimation benefits do not apply. We observe autobalancing benefits for fixed P but not for fixed N when Σ is known, which implies an interaction between autobalancing and the constraints. Thus, the conditions for autobalancing warrant a more detailed examination. First, what are the designs in which autobalancing significantly helps? Second, when is the effect of autobalancing large? To answer these questions, this section conducts a large-scale simulation study of the MNL (known Σ), focusing on the determinant criterion throughout.

We use the closed-form expression for the Fisher information for the MNL to examine a greater range of DCA designs than was feasible with the MNP and HB-MNP. We consider the following design variables potential moderators of the autobalancing effect: total number of brands (K), proportion of dominated brands (denoted D), number of tasks (N), number of nonbrand attributes (B = dimension of β), and strength of the effects (absolute magnitude of α and β). Given these variables, we ask two questions of managerial relevance: First, how does the size of the optimal data set depend on the design variables? Second, how does the magnitude of the accuracy improvement from using the optimal subset size depend on the design variables?

To manipulate the strength of the effects and the proportion of dominated brands, we fix a proportion D of the scalar elements of α to –c and fix the remainder to c, where c is a constant. To simultaneously control the strength of the effects of nonbrand attributes, we set the scalar elements of β to evenly spaced numbers between –c and c for the same constant. Specifically, we set $\beta_1 = -c$, $\beta_2 = -c + 2c/(B - 1)$, …, $\beta_B = c$. Therefore, the constant c can be interpreted as effect strength relative to the standard deviation of the Gumbel utility shock, which is $\pi/\sqrt{6} = 1.28$. In this sense, c = .5 are "weak" effects, c = 1 are "medium" effects, c = 2 are "strong" effects, and so on. Table 4 gives the values of the parameters we consider in a full factorial study consisting of 3 × 2 × 3 × 4 × (3 + 7 + 11 + 15 + 19) = 3960 conditions.

For every combination of the design variables, we draw a random X matrix with binary nonbrand attributes, vary the size of random subsets S = 2, …, K, and compute the log-determinant of the Fisher information matrix. We use the log-transformation because the determinants are highly skewed. To normalize across different simulations, we then consider the ratio of the resultant log-information in random sets of size S to the log-information in the full-dimensional sets of size K. The information matrix depends on the particular realization of the nonbrand attributes X and on the particular subsets $\{C_n\}$ of size S. To assess this variability, we generate 100 X draws for every design setting and draw different random subsets every draw. Therefore, our study considers 396,000 different data sets. The final measure we report and analyze is the expected relative log-information:

Table 4

MNL SIMULATION STUDY: DESIGN AND A SUMMARY OF RESULTS

Design Variable	*Values Considered*	*Effect on Benefit of Subsets*
Number of tasks N	500, 1000, 3000	None
Proportion of dominated brands D	.5, .75	+
Number of nonbrand two-level attributes B	3, 5, 10	–
Strength of effects c	.5, 1, 2, 3	+
Number of brands K	4, 8, 12, 16, 20	+
Subset size S	2, 3, …, K	

$$\text{(8)}\qquad \text{Expected relative log Info}\left(S/K \mid K, c, B, N\right) = E_X\left[\frac{\log\det I_K^S\left(K, c, B, D, N \mid X\right)}{\log\det I_K^K\left(K, c, B, D, N \mid X\right)}\right].$$

Results

To answer the first question about the optimal subset size, we consider the 360 conditions arising from the factorial design using all the variables in Table 4, except for S. For each of the 360 conditions and each of the 100 simulated data sets, we find the optimal subset size S*/K. As expected from theory and the MNP simulations, strict subsets are not always optimal; we find that S*/K < 1 in only 23% of the cases. In the remaining 77%, the designer is best off with full-dimensional choice sets. The number of brands K is perhaps the most important moderator of autobalancing: We find that, at least within our range of the design variables, random subsets are never optimal for K = 4, and the percentage of cases with optimal subsets then rises with K to 11% (K = 8), 25% (K = 12), 35% (K = 16), and finally 43% (K = 20). Given this observation that optimality of strict subsets is not universal, we now turn to a systematic investigation of which design situations make strict subsets optimal.

To analyze the first-order effects of the design variables on the optimal subset size, we run a simple linear regression of S*/K on design variable dummies (for the output, see Table W2 in the Web Appendix at http://www.marketingpower.com/jmrdec09). The regression identifies three key variables: Subsets are beneficial when there are more brands (K higher), when there are fewer nonbrand attributes (B lower), and when the effects are stronger (c higher). As we expected from Examples 2 and 3, the effect strength variable c interacts positively with the ratio D of dominated brands. The main effects and one interaction (c × D) explain roughly half of the variation in optimal subset size (R^2 = .50). Approximately one-tenth of the remaining unexplained variation is due to variation across the 100 data sets within each condition (when we average the S*/K across the 100 data sets in each condition between regressing, the R-square of the resulting regression is .54). Finally, approximately another half of the remaining variation after pooling across the 100 data draws can be attributed to our pooling across different K (running the pooled regressions separately within each K yields an R-square of approximately .7). These results suggest that the weak autobalancing benefit in Tables 1 and 3 for the constant number of tasks and known Σ case is due to not having a sufficiently large K and not having strong-enough effects for autobalancing to make strict subsets optimal.

To answer our second question regarding effect size, we regress the relative log-information on subset size (expressed as a proportion S/K), squared subset size, and the design variables. Table W2 in the Web Appendix again shows the results (http://www.marketingpower.com/jmrdec09): Positive signs indicate positive benefit of subsets, and vice versa. It is immediately clear that the most important determinants of effect size are the same as those in the optimal-size regression we discussed previously. Moreover, the magnitude of information gain/loss clearly depends on the number of brands, which generally increases as the number of brands (K) increases. For example, having ten versus three nonbrand attributes results in an average log-information loss of .6% with K = 4 and as much as 7% with K = 20. The information gain due to stronger effects exhibits a complicated three-way interaction with the percentage of dominated brands and the total number of brands: The largest information gain seems to be available when effects are strong, brands are many, and the percentage of dominated brands is higher. Figure 2 illustrates this interaction.

The results here suggest a few rules of thumb to designers of choice experiments who intend to use the MNL in analyzing their data. We find that relatively smaller subsets are more informative than full sets under the conservative resource constraint when the following apply:

- There are many brands to evaluate (eight or more),
- The study considers fewer nonbrand attributes (fewer than five),
- The effects are likely to be strong (utility partworths of absolute magnitude of two or greater), and
- The effects are likely to be strong, and a majority of the brands are dominated.

Unlike the other two benefits of subsets, the autobalancing benefit occurs only in specific situations. One of the contributions of this article is to characterize when the benefit operates. The optimal subset size depends on the specific values of the design variables, and in general, it stays above .8K. Although a few extreme conditions make even smaller subsets optimal (for the situation K = 20, c = 3, and D = .75, for which 6 = .3K is the optimal size, see Figure 2), we find that very small subsets, like pairs, are not optimal. Finally, the slope of the relative accuracy improvement is asymmetric around the optimum value: In general, using a slightly larger-than-optimal subset results in a much smaller information loss than using a slightly smaller-than-optimal subset (for an illustration, see Figure 2). Therefore, somewhat large subsets are the conservative choice.

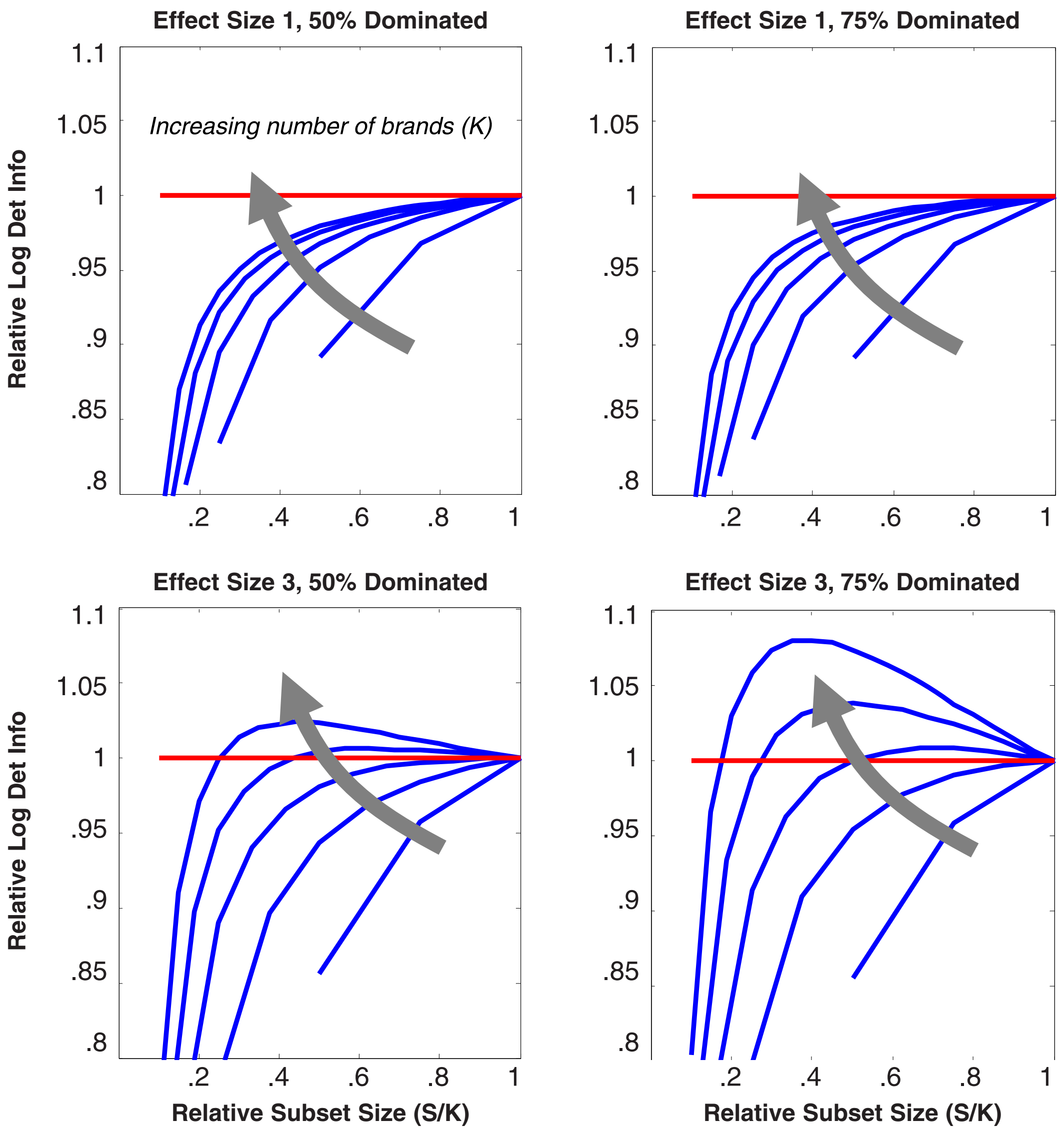
DISCUSSION

We demonstrate that choices from random subsets of brands can contain more information about demand than choices from all brands under study. Three different statistical reasons contribute to the findings: Subsets allow more choice tasks per respondent, subsets improve the estimation of correlations between random utilities, and subsets can have more balanced utilities than full sets. Which subset size a survey designer should use depends on the relationship between size and respondents' cognitive difficulty, on the particular model used to estimate demand, and on properties of the demand itself. Depending on these moderating conditions, optimal subsets can range from pairs to the complete sets of all brands under study, and the improvements from using the optimal subset size can be large. Therefore, subset size should be an important choice-based conjoint design variable. This article provides the theoretical basis for understanding why and when subsets can help and gives several rules of thumb about the best subset sizes to use.

The most robust benefit of subsets arises from the notion that choices from smaller choice sets are quicker and easier to make, so surveys using smaller subsets can use more choice tasks per respondent, given a constraint on the total number of profiles. We find that a choice experiment designer who faces this resource constraint will always find strict subsets to be beneficial, but the optimal subset size and the magnitude of the benefit depend on the moderators introduced in the previous paragraph. Pairs are optimal for the homogeneous MNL with only brand intercepts, whereas this corner solution finding does not generalize to more complex settings. The HB-MNP model calibrated on real-world conjoint data provides probably the most practically relevant results. Of the five brands in the study plus the outside option, the optimal subset size is a triple or a quadruple depending on the measure. In our studies, the brand equity estimate—that is, the standard deviation of the change in share predicted as a result of setting a product's brand partworth to zero—benefits the most from subsets with a 9%–15% decrease when the subsets used are triples versus when they are the full-set sextuples.

Even when the number of choice tasks is held constant, so subsets require fewer product profiles, subsets offer benefits: improved variance estimation and automatic utility balancing for a brand with low-brand partworths. Subsets act as exclusion restrictions in the data, and so choices from subsets can better separate brand intercepts from brand correlations. The result is easier identification and more precise estimation of the covariance structure, which in turn can increase posterior precision of most statistics of interest. As would be expected, measures such as the aforementioned brand equity benefit more than measures not critically related to specific brand-level estimates, such as overall holdout performance or expected lift due to a nonbrand attribute. Homogeneous MNL and MNP have considerable gains, with 20%–40% increased accuracy on some measures while using 30%–50% fewer profiles. Heterogeneity in response blunts this benefit, and we find

Figure 2
BENEFIT OF SUBSETS IN THE MNL AS A FUNCTION OF EFFECT SIZE, PROPORTION OF DOMINATED BRANDS, AND THE OVERALL NUMBER OF BRANDS



Notes: Each line corresponds to a model with a different number of brands, plotting average across remaining treatments for all relative subset sizes. The gray arrows indicate increasing the number of brands from 4 to 20. Det info = determinant of Fisher information.

that HB-MNP models usually gain less from subsets than the homogeneous MNP.

Autobalancing operates when the brands are unbalanced in the utility function. When that is the case, the dominated brands are rarely chosen from the full set of brands, and the survey designer learns little about them. A well-known solution is balancing the choice sets, for example, by reducing the prices of the dominated brands. However, such balancing requires knowledge of the very parameters the survey is supposed to estimate. In contrast, random subsets approximate balancing automatically because the relatively dominant brand will be excluded from some tasks. Our theoretical analysis and simulations demonstrate this benefit when the degree of dominance is sufficiently large. Four rules of thumb from a large scale simulation are as follows: Strict subsets are beneficial when (1) more than eight brands are in the study, (2) there are fewer than five nonbrand attributes to consider, (3) the utility partworths of the dominating brands are large, and (4) the proportion of dominated brands is large. The autobalancing benefit can operate in large-scale discrete choice analyses, especially when their focus is on niche marketing and many relatively small brands competing with a dominant mass-market competitor.

The statistical benefits of subsets are separate and additional to the better-known behavioral benefits. When faced with large choice sets, respondents switch from careful compensatory assessment to simplifying heuristics (Payne, Bettman, and Johnson 1988). As a result, the model errors are smaller with smaller subsets (DeShazo and Fermo

2002). More laboratory work is needed to estimate the size of these behavioral benefits from subsets and to assess their joint impact on and possible interactions with the statistical benefits we propose.

We also study the posterior precision of difference-in-probability measures, such as brand equity and sale lift—that is, the effect of a binary attribute. Such counterfactual questions are often the reason a conjoint analysis is conducted, so understanding their statistical properties better is managerially relevant. Notably, we find that the situations with the most precise estimates of the level of probability (i.e., the best holdout performance) do not correspond to the situations with the most precise difference-in-probability estimates. In terms of subset size, we find that relatively smaller subsets are usually optimal for the latter measures. Fully understanding this discrepancy and determining the "correct" measure for judging estimation precision remains a goal for future work.

APPENDIX: PROOFS OF PROPOSITIONS AND DETAILS OF EXAMPLES

Proof of Lemma 1

Consider the (α, α) submatrix of $-H_n(\alpha, \beta|S)$ in Equation 3. Because the subsets are random and equally likely, the probability of a brand being present in C_n is S/K, and the probability of any given pair of brands is $(S/K)[(S-1)/(K-1)]$. Because p_k^S does not depend on n, the total information $I_K^S(\alpha)$ in the design is as follows:

$$I_K^S(\alpha) = \sum_n -H_n^S(\alpha \mid C_n, S<K) = N\left(\frac{S}{K}\right)$$

$$\begin{bmatrix} p_1^S(1-p_1^S) & -\left(\frac{S-1}{K-1}\right)p_1^S p_j^S & -\left(\frac{S-1}{K-1}\right)p_1^S p_{K-1}^S \\ \cdots & p_j^S(1-p_j^S) & -\left(\frac{S-1}{K-1}\right)p_j^S p_{K-1}^S \\ \cdots & \cdots & p_{K-1}^S(1-p_{K-1}^S) \end{bmatrix}.$$

Then, $p_k^S = 1/S$ implies that the determinant of $I_K^S(\alpha)$ can be computed as follows:

$$\text{(A1)} \qquad \det\left[I_K^S(\alpha)\right] = N^{K-1}$$

$$\det\begin{bmatrix} \left(\frac{S}{K}\right)\left(\frac{S-1}{S^2}\right) & \cdots & -\left(\frac{S}{K}\right)\left(\frac{S-1}{K-1}\right)\left(\frac{1}{S^2}\right) \\ \cdots & \left(\frac{S}{K}\right)\left(\frac{S-1}{S^2}\right) & \cdots \\ \cdots & \cdots & \left(\frac{S}{K}\right)\left(\frac{S-1}{S^2}\right) \end{bmatrix}.$$

Factoring $(S-1)/(KS)$ from every row gives the following:

$$\det\left[I_K^S(\alpha)\right] = N^{K-1}\left(\frac{S-1}{KS}\right)^{K-1} \det\begin{bmatrix} 1 & \cdots & -\frac{1}{K-1} \\ \cdots & 1 & \cdots \\ \cdots & \cdots & 1 \end{bmatrix}.$$

Now, we add all other columns to the first column (the determinant is invariant to this operation). The first column becomes $1/(K-1)$ in every row, so we subtract the first row from every other row:

$$\begin{vmatrix} 1 & \cdots & -\frac{1}{K-1} \\ \cdots & 1 & \cdots \\ \cdots & \cdots & 1 \end{vmatrix} = \begin{vmatrix} \frac{1}{K-1} & -\frac{1}{K-1} & -\frac{1}{K-1} \\ \frac{1}{K-1} & 1 & -\frac{1}{K-1} \\ \frac{1}{K-1} & -\frac{1}{K-1} & 1 \end{vmatrix}$$

$$= \begin{vmatrix} \frac{1}{K-1} & -\frac{1}{K-1} & -\frac{1}{K-1} \\ 0 & \frac{K}{K-1} & 0 \\ 0 & 0 & \frac{K}{K-1} \end{vmatrix} = \frac{1}{K}\left(\frac{K}{K-1}\right)^{K-1}.$$

Therefore,

$$\det\left[I_K^S(\alpha)\right] = N^{K-1}\frac{1}{K}\left[\frac{(S-1)}{S(K-1)}\right]^{K-1}.$$

The trace result follows immediately by adding the diagonal in Equation A1. Q.E.D. Lemma 1.

Technical Details of Example 1

The full set {A, B, C} reveals only two nonredundant pieces of information: s_A = share of A, and s_B = share of B. Both are complicated nonlinear functions of all parameters $\{\alpha_A, \alpha_B, \rho\}$:

$$s_A(\alpha_A, \alpha_B, \rho) =$$

$$\int_{-\alpha_A}^{\infty} \int_{-\infty}^{\alpha_A - \alpha_B + \varepsilon_A} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{\varepsilon_A^2 - 2\rho\varepsilon_A\varepsilon_B + \varepsilon_B^2}{2}\right) d\varepsilon_A d\varepsilon_B.$$

The model cannot be identified, because there are more unknown parameters (three) than observational equations (two). Now, consider choices from subsets (pairs). Pairs reveal three pieces of information: $s_{A|AB}$ = share of A from {A,B}, $s_{A|AC}$ = share of A from {A,C}, and $s_{B|BC}$ = share of B from {B, C}. The model is now identified as long as $\alpha \neq \beta$: First, $s_{A|AC}$ pins down α because $s_{A|AC}(\alpha_A, \alpha_B, \rho) = s_{A|AC}(\alpha_A) = \Phi(-\alpha_A)$, where Φ is the standard normal cumulative distribution function. Second, $s_{B|BC}$ pins down β analogously because $s_{B|BC} = \Phi(-\alpha_B)$. Third, the model is identified if the third share $s_{A|AB}$ identifies ρ conditional on α and β. Suppose without loss of generality that $\alpha_A \geq \alpha_B$ (i.e., A weakly dominates B). From $\varepsilon_B|\varepsilon_A \sim \rho\varepsilon_A + N(0, 1-\rho^2)$, we rewrite

$$s_{A|AB} = \Pr(\alpha_A - \alpha_B > \varepsilon_B - \varepsilon_A)$$

$$= \Pr\left[\alpha_A - \alpha_B > \eta\sqrt{1-\rho^2} - (1-\rho)\varepsilon\right],$$

where ε and η are uncorrelated standard normal random variables. The sum of two uncorrelated normals is a normal with variance equal to the sum of the two variances. Thus,

$$s_{A|AB} = \Phi\left[\frac{\alpha_A - \alpha_B}{\sqrt{2(1-\rho)}}\right].$$

When $\alpha_A = \alpha_B$, this implies that $s_{A|AB} = 1/2$ for every ρ, and so ρ is not identified. Conversely, when $\alpha_A > \alpha_B$, $s_{A|AB}$ increases monotonically in ρ because

$$\frac{d}{dr}\left[\frac{\alpha_A - \alpha_B}{\sqrt{2(1-\rho)}}\right] > 0$$

and Φ is increasing. Therefore, by symmetry, $s_{A|AB}(\rho)$ is invertible whenever $\alpha_A \neq \alpha_B$. Q.E.D. Example 1.

Technical Details of Example 2

Claim. Let K = 3, and fix $\alpha_3 = 0$, where 3 is the dominant brand. For simplicity, we assume that 1 and 2 have the same negative value, $\alpha < -5/2$. Then, N random pairs are more Fisher-informative than N triples.

Proof. Consider an equal mixture of subsets A = {1, 2}, B = {1, 3}, and C = {2, 3}, with the respective choice-probabilities p_k^A, p_k^B, and p_k^C. Thus,

$$p_1^A = p_2^A = \frac{1}{2}, p_1^B = p_2^C = \frac{\exp(\alpha)}{1+\exp(\alpha)} \equiv p, p_3^B = p_3^C = 1 - p.$$

Let the probability of choice from the full set F = {1, 2, 3} be

$$q \equiv p_1^F = p_2^F = \frac{\exp(\alpha)}{1 + 2\exp(\alpha)}.$$

In this design, the single-task Hessians H_n^2 $(\alpha \mid C_n)$ are as follows:

$$-H_n^2(\alpha \mid A) = \begin{bmatrix} p_1^A(1-p_1^A) & -p_1^A p_2^A \\ -p_1^A p_2^A & p_2^A(1-p_2^A) \end{bmatrix} = \left(\frac{1}{4}\right)\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

$$-H_n^2(\alpha \mid B) = \begin{bmatrix} p_1^B(1-p_1^B) & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} p(1-p) & 0 \\ 0 & 0 \end{bmatrix}, \text{and}$$

$$-H_n^2(\alpha \mid C) = \begin{bmatrix} 0 & 0 \\ 0 & p_2^C(1-p_2^C) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & p(1-p) \end{bmatrix}.$$

Therefore,

$$I_3^2(\alpha) = \frac{N}{3}\begin{bmatrix} p_1^A(1-p_1^A) + p_1^B(1-p_1^B) & -p_1^A p_2^A \\ -p_1^A p_2^A & p_2^A(1-p_2^A) + p_2^C(1-p_2^C) \end{bmatrix}$$

$$\Rightarrow \det(I_3^2) = \frac{N^2}{9}\left[p_1^A p_2^A p_2^C p_3^C + p_1^A p_2^A p_1^B p_3^B + p_1^B p_3^B p_2^C p_3^C\right]$$

$$= N^2\frac{p(1-p)}{9}\left[\frac{1}{2} + p(1-p)\right].$$

We compare this information with the full-set design's $\det(I^3) = N^2 p_1^F p_2^F p_3^F = N^2 q^2(1 - 2q)$. We can show that for $\alpha < -5/2$, $\det(-H^3) < \det(-H^2)$. To understand why a low-enough α makes the random subsets more informative, let $z = \exp(\alpha)$, so the $\det(-H^3) < \det(-H^2)$ inequality holds when $1 - 8z - 35z^2 - 46z^3 - 28z^4 - 10z^5 > 0$. This is clearly satisfied at z = 0, and continuity implies that it is satisfied for a small-enough z > 0 as well. The left-hand-side polynomial is decreasing in z, so it has only one positive real root of approximately

$$\exp\left(-\frac{5}{2}\right).$$

The root is not analytically tractable.

Now, consider the trace summary of the information matrix. With the equal mixture of subsets,

$$\text{trace}(I_3^2) = \frac{N}{3}\left[\frac{1}{2} + 2p(1-p)\right].$$

With full sets (triples), $\text{trace}(I_3^3) = N[2q(1 - q)]$. Let $z = \exp(\alpha)$, and note that $\text{trace}(I_3^2) > \text{trace}(I_3^3)$ whenever $1 - 2z + 5z^2 + 16z^3 + 4z^4 > 0$. This inequality is satisfied for all z > 0 and, thus, for all α because its LHS is bounded below by $1 - 2z + 2z^2 = (1 - z)^2 + z^2 > z^2 > 0$. Therefore, based on the trace criterion, N random pairs are preferred to N triples for all values of α. Q.E.D. Example 2.

REFERENCES

Allenby, Greg and Peter Lenk (1994), "Modeling Household Purchase Behavior with Logistic Normal Regression," *Journal of the American Statistical Association*, 83 (428), 1218–31.

——— and ——— (1995), "Reassessing Brand Loyalty, Price Sensitivity, and Merchandizing Effects on Consumer Brand Choice," *Journal of Business and Economic Statistics*, 13 (3), 281–90.

Arora, Neeraj and Joel Huber (2001), "Improving Parameter Estimates and Model Prediction by Aggregate Customization in Choice Experiments," *Journal of Consumer Research*, 28 (September), 273–83.

Ben-Akiva, Moshe and Steven R. Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Bettman, James R., E.J. Johnson, and J.W. Payne (1990), "A Componential Analysis of Cognitive Effort in Choice," *Organizational Behavior and Human Decision Processes*, 45 (1), 111–39.

Chamberlain, G. (1986), "Asymptotic Efficiency in Semiparametric Models with Censoring," *Journal of Econometrics*, 32 (1), 189–218.

Currim, Imran S. (1982), "Predictive Testing of Consumer Choice Models Not Subject to Independence of Irrelevant Alternatives," *Journal of Marketing Research*, 19 (May), 208–222.

DeShazo, J.R. and G. Fermo (2002), "Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency," *Journal of Environmental Economics and Management*, 43 (3), 360–85.

Elrod, Terry and Michael P. Keane (1995), "A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data," *Journal of Marketing Research*, 32 (February), 1–16.

Gilbride, Tim and Greg Allenby (2006), “Estimating Heterogeneous EBA and Economic Screening Rule Choice Models,” *Marketing Science*, 25 (5), 494–509.

Goldfarb, Avi, Qiang Lu, and Sridhar Moorthy (2009), “Measuring Brand Value in an Equilibrium Framework,” *Marketing Science*, 27 (1), 69–86.

Huber, Joel and Christopher Puto (1983), “Market Boundaries and Product Choice: Illustrating Attraction and Substitution Effects,” *Journal of Consumer Research*, 10 (June), 31–44.

——— and Klaus Zwerina (1996), “The Importance of Utility Balance in Efficient Choice Designs,” *Journal of Marketing Research*, 33 (August), 307–317.

Kamakura, Wagner A. and Gary J. Russell (1993), “Measuring Brand Value with Scanner Data,” *International Journal of Research in Marketing*, 10 (March), 9–22.

Keane, Michael P. (1992), “A Note on Identification in the Multinomial Probit Model,” *Journal of Business & Economic Statistics*, 10 (2), 193–200.

Kuhfeld, Warren, Randall D. Tobias, and Mark Garratt (1994), “Efficient Experimental Design with Marketing Research Applications,” *Journal of Marketing Research*, 31 (November), 545–57.

Lenk, Peter, Wayne DeSarbo, Paul Green, and M. Young (1996), “Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs,” *Marketing Science*, 15 (2), 173–91.

Louviere, Jordan J., Deborah J. Street, and Leonie Burgess (2003), “A 20+ Years Retrospective on Choice Experiments,” in *Marketing Research and Modeling: Progress and Prospects*, Yoram Wind and Paul E. Green, eds. New York: Kluwer Academic Press, 201–214.

McCulloch, Robert and Peter Rossi (1994), “An Exact Likelihood Analysis of the Multinomial Probit Model,” *Journal of Econometrics*, 64 (1), 207–240.

Orme, Bryan (2005), *Getting Started with Conjoint Analysis: Strategies for Product Design and Product Research*. Madison: Research Publishers.

Payne, J.W., J.R. Bettman, and Eric Johnson (1988), “Adaptive Strategy Selection in Decision Making,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14 (1), 534–52.

Pinnell, J. and S. Englert (1997), “The Number of Choice Alternatives in Discrete Choice Modeling,” in *1997 Sawtooth Software Conference Proceedings*. Ketchum, ID: Sawtooth Software, 121–53.

Rossi, Peter, Greg Allenby, and Robert McCulloch (2005), *Bayesian Statistics and Marketing*. Hoboken, NJ: John Wiley & Sons.

Sandor, Zsolt and Michel Wedel (2001), “Designing Conjoint Choice Experiments Using Managers’ Prior Beliefs,” *Journal of Marketing Research*, 38 (November), 430–44.

Tversky, Amos and Itamar Simonson (1993), “Context-Dependent Preferences,” *Management Science*, 39 (October), 1179–89.

Zeithammer, Robert and Peter Lenk (2006), “Bayesian Estimation of Multivariate-Normal Models When Dimensions Are Absent,” *Quantitative Marketing and Economics*, 4 (3), 241–65.