Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs

Peter J. Lenk • Wayne S. DeSarbo • Paul E. Green • Martin R. Young

The University of Michigan
Pennsylvania State University
University of Pennsylvania
The University of Michigan

Abstract

The drive to satisfy customers in narrowly defined market segments has led firms to offer wider arrays of products and services. Delivering products and services with the appropriate mix of features for these highly fragmented market segments requires understanding the value that customers place on these features. Conjoint analysis endeavors to unravel the value, or partworths, that customers place on the product or service's attributes from experimental subjects' evaluation of profiles based on hypothetical products or services. When the goal is to estimate the heterogeneity in the customers' partworths, traditional estimation methods, such as least squares, require each subject to respond to more profiles than product attributes, resulting in lengthy questionnaires for complex, multiattributed products or services. Long questionnaires pose practical and theoretical problems. Response rates tend to decrease with increasing questionnaire length, and more importantly, academic evidence indicates that long questionnaires may induce response biases.

The problems associated with long questionnaires call for experimental designs and estimation methods that recover the heterogeneity in the partworths with shorter questionnaires. Unlike more popular estimation methods, Hierarchical Bayes (HB) random effects models do not require that individual-level design matrices be of full rank, which leads to the possibility of using fewer profiles per subject than currently used. Can this theoretical possibility be practically implemented?

This paper tests this conjecture with empirical studies and mathematical analysis. The random effects model in the paper describes the heterogeneity in subject-level partworths or regression coefficients with a linear model that can include subject-level covariates. In addition, the error variances are specific to the subjects, thus allowing for the differential use of the measurement scale by different subjects.

In the empirical study, subjects' responses to a full profile design are randomly deleted to test the performance of HB methods with declining sample sizes. These simple experiments indicate that HB methods can recover heterogeneity and estimate individual-level partworths, even when individual-level least squares estimators do not exist due to insufficient degrees of freedom.

Motivated by these empirical studies, the paper analytically investigates the trade-off between the number of profiles per subject and the number of subjects on the statistical accuracy of the estimators that describe the partworth heterogeneity. The paper considers two experimental designs: each subject receives the same set of profiles, and subjects receive different blocks of a fractional factorial design. In the first case, the optimal design, subject to a budget constraint, uses more subjects and fewer profiles per subject when the ratio of unexplained, partworth heterogeneity to unexplained response variance is large. In the second case, one can maintain a given level of estimation accuracy as the number of profiles per subject decreases by increasing the number of subjects assigned to each block.

These results provide marketing researchers the option of using shorter questionnaires for complex products or services. The analysis assumes that response quality is independent of questionnaire length and does not address the impact of design factors on response quality. If response quality and questionnaire length were, in fact, unrelated, then marketing researchers would still find the paper's results useful in improving the efficiency of their conjoint designs. However, if response quality were to decline with questionnaire length, as the preponderance of academic research indicates, then the option to use shorter questionnaires would become even more valuable.

(*Consumer Preferences; Multi-attributed Models; Consumer Research*)

0732-2399/96/1502/0173$01.25

1. Introduction

Over the past three decades, conjoint analysis has evolved as the primary set of techniques employed by both academics and practitioners of marketing research for measuring consumer tradeoffs among multiattributed products and services. Originally devised by Green and Rao (1971), this decompositional procedure utilizes overall evaluations of typically hypothetical products or services defined on experimentally manipulated attributes, and derives quantitative values for attribute levels from these evaluations. Green and Krieger (1994a), Green and Srinivasan (1978, 1990), and Louviere (1988) have provided summaries of the major phases and associated methodological decisions employed in the execution of a typical marketing conjoint study (see also Rao 1977). Cattin and Wittink (1982), Wittink and Cattin (1989), and Wittink et al. (1994) have documented the extensive use of conjoint analyses involving a wide assortment of commercial applications.

In the full profile method of data collection, respondents evaluate hypothetical product or service profiles described by their attribute levels, which are specified from an experimental design. Frequently, studies employing validation with hold-out profiles or market share choice simulations require individual consumer-level responses, which are inputs to the simulators. To obtain subject-level estimates, subjects often must rank, rate, or choose among a large number of profiles, especially if there is a large number of attributes and levels, if traditional estimation techniques are used. Even for the most involved respondents, the task is often characterized as excessively demanding, time consuming, boring, and frustrating (Malhotra 1986). The burden of data collection is further accentuated in studies involving complex interaction terms.

Efficient experimental designs are particularly important in conjoint studies for financial reasons as well as maintaining response quality. A large body of research in psychology and marketing indicates that questionnaire length, complexity, relevance, and interest affect response rates and can potentially introduce response biases. The preponderance of evidence indicates that, all else being equal, longer questionnaires have lower response rates than shorter ones (Adams and Gale 1982, Bean and Roszkowski 1995, Brown et al. 1989, Dillman 1991, Dillman et al. 1993, Heberline and Baumgartner 1978, Lockhart 1990, and Roszkowski and Bean 1990). Berdie (1989) found that response rates and response bias are related. Low response rates can be partly ameliorated in conjoint studies by using paid subjects (Biner 1988, Biner and Barton 1990, Biner and Kidd 1994, Brennan 1992, Gunn and Rhodes 1981, Hansen 1980, Huck and Gleason 1974, Paolillo and Lorenzi 1984, and Yu and Cooper 1983).

A more subtle challenge arises due to response bias or response quality. Kraut et al. (1975) showed an increasing tendency on the part of subjects to use the modal response category instead of the extreme ones in later parts of a lengthy questionnaire. Herzog and Bachman (1981) confirmed this result and dubbed it, "straight-line stereotyping." In a different venue, Dong (1983) investigated the quality of subjects' responses in judging triads and found more violations of the triangle inequality occurred toward the end of a long experiment.

One response to the problems of using a long questionnaire has been to select a few attributes that the researcher believes a priori most salient to the decision process. This culling of attributes presents its own problems. Jacoby et al. (1977), Payne (1976), and Wright (1974) found that subjects faced with multiattribute decisions use simplifying heuristics that do not simultaneously consider the value of all options on all attributes. The implication for conjoint studies is that the profiles presented to the subject should be congruent, with respect to their attributes, to those in the marketplace, or else the decision process during the conjoint study may substantially differ from that during the actual purchase occasion. Therefore, attempts to restrict the conjoint study to relatively few attributes risk biasing the subjects' evaluations by directing their attention to an array of attributes that they may not focus on during the actual purchase.

The complexity of the task can affect a subjects decision strategy, but what of the quality of their responses? Jacoby et al. (1974) and Jacoby et al. (1974) found that subjects in conjoint studies made "poorer purchase decisions" as the information load, defined to be the number of brands times the number of dimensions, increases. These papers put forth the hypothesis that decision quality is an inverted "U" shape with respect to the amount of information. Too much information, called "information overload," results in less effective decisions. This hypothesis is difficult to address due to the difficulty of measuring the quality of a subject's preferences and the amount of information, and has not been without controversy (see Jacoby 1977 and 1984, Malhotra 1982 and 1984, Malhotra et al. 1982, Summers 1974, and Wilkie 1974). Keller and Staelin (1987) also examined information overload and found that decision effectiveness tends to increase with increases of information quality and with decreases in information quantity. (Also, see the commentary of Meyer and Johnson 1989 and Keller and Staelin's 1989 response.) In a more recent study, Helgeson and Ursic (1993) confirmed the decline in decision accuracy with the complexity of the information load.

While the information overload hypothesis has not been conclusively answered due to measurement problems, the preponderance of evidence indicates that burdening the consumer with complex, irrelevant, and lengthy tasks may negatively affect their response rate and may introduce biases in their responses. To the extent that marketing researchers can reduce potential boredom, fatigue, and disinterest from conjoint studies (cf. Malhotra 1986), crisper partworth estimates should be obtained that are more representative of the desired populations. Jedidi et al. (1995) have also mentioned this issue and have recommended the use of shorter, more concise conjoint response tasks.

Researchers have responded in various ways to the implicit tradeoff between response burden, as measured by the length and complexity of the conjoint study, and accurate estimation of individual differences in partworths. The work of Allenby et al. (1995), Srinivasn et al. (1983), and van der Lans and Heiser (1992) address one aspect of data quality by incorporating either constraints derived from the researcher or respondents on the ordering of partworths. Individually based hybrid models (Green and Krieger 1994b) and Adaptive Conjoint Analysis (Johnson 1987) address the problem by collecting relatively easy-to-obtain, self-explicated data prior to the collection of a set of full or partial response. While these procedures do not necessarily force order constraints on the partworths, in practice the self-explicated data constitute a significant part of the total input and, hence, exert a strong influence on the final partworths (Green et al. 1991). Optimal scaling (Hagerty 1985), clustering and latent class approaches (Kamakura 1988, DeSarbo et al. 1989, Wedel and Steenkamp 1989, DeSarbo et al. 1992), and empirical Bayes models (Green et al. 1993) attempt, in various ways, to pool data across individuals.

This paper continues the investigation of efficient conjoint design and estimation by empirically and analytically exploring the feasibility of estimating the heterogeneity of the partworths when fewer profiles per subject are used as compared to more traditional methods. Random effects models provide a natural method to describe the heterogeneity in the individual level partworths. The model assumes that a subject's metric response to a full profile of attributes is a linear function of these attributes. The individual-level partworths vary across the population according to probability distributions that can depend on covariates such as demographic variables or prior usage. Also, we modify the standard model, which Allenby and Ginter (1995) use, by introducing heterogeneity in the error variances. By assuming individual-specific intercepts and error variances, the methodology automatically adjusts the estimators for the subjects' differential use of the measurement scale in both location and spread.

Two goals of random effects models are to infer individual-level parameters and the distributions that describes their heterogeneity. Estimation accuracy depends on the number of profiles per subject and the number of subjects in the study. At one extreme, one subject's evaluation of a large number of profiles leads to accurate estimation of that subject's partworths, but provides only one data point for estimating the distribution of partworths across the population. If partworths are completely homogeneous, then only one subject is needed in the experiment. At another extreme, both individual partworths and their heterogeneity can be accurately estimated when many subjects respond to a large number of profiles. Unfortunately, such an experiment is often too expensive. In many cases, subjects may be unwilling to respond to the full set of profiles, or there may be response biases, as previously documented. In this situation, it would be desirable to elicit responses to a reduced set of profiles.

This paper investigates the tradeoff between the number of profiles per subject and the number of subjects in accurately representing customer heterogeneity. We maintain that it is possible to obtain comparable levels of accuracy in estimating partworth heterogeneity when different groups of subjects respond to different subsets of profiles, each subset having fewer profiles than a complete design. The tradeoff is that more subjects are needed for the study.

The model is estimated in a hierarchical Bayes (HB) framework. Hill (1965) originally presented the Bayesian analysis of random effects models. Lindley and Smith (1972) and Smith (1973) describe the HB analysis of linear models. Berger (1985) provides a review of HB models and their analysis. Recent applications of HB models to marketing include new product diffusion (Lenk and Rao 1990), coupon redemptions (Lenk 1992), and brand choice (Allenby and Lenk 1994 and 1995).

The next section of the paper presents the proposed model. Section 3 presents two empirical examples. The first example uses synthetic data with known structure, and the second reports the results of a conjoint study of personal computers. The examples test the recovery of partworth heterogeneity as profiles are randomly deleted and demonstrate that of HB models can be used even when the number of partworths exceed the number of profiles per subject. Note, however, that the examples do not investigate the information overload hypothesis. Section 4 analytically describes the dependency of estimation accuracy on the number of subjects and the number of profiles per subject and recommends two optimal designs. Finally, the Discussion delineates some limitations of the hierarchical Bayes methodology, as well as directions for future research.

2. Hierarchical Bayes Conjoint Analysis

The random effects model for conjoint analysis describes the variation in a subject's responses and the variation in the subjects' partworths over the population:

$$Y_i = X_i\beta_i + \epsilon_i \quad \text{for } i = 1, \ldots, n, \tag{1}$$

$$\beta_i = \Theta z_i + \delta_i \quad \text{for } i = 1, \ldots, n. \tag{2}$$

In Equation (1), Y_i is a vector of m_i metric responses for subject i to the profiles described by a given design matrix X_i, and β_i is the p-dimensional vector of regression coefficients or partworths for subject i. Equation (2) describes the heterogeneity of individual-level partworths via a multivariate regression model with q-dimensional covariates, z_i, and a p by q matrix of regression coefficients, Θ. In the simplest case, z_i is equal to 1 for all i, and Θ is the mean vector for the partworths.

The error terms $\{\epsilon_i\}$ and $\{\delta_i\}$ in Equations (1) and (2) are assumed to be mutually independent and from multivariate normal distributions with zero means and covariance matrices $\{\sigma_i^2 I\}$ and Λ, respectively: ϵ_i is $N_{m_i}(0, \sigma_i^2 I)$ and δ_i is $N_p(0, \Lambda)$ where I is the identity matrix, and Λ is a $p \times p$ positive definite matrix. In addition to heterogeneity in the partworths, we assume that the error variances $\{\sigma_i^2\}$ form a random sample from an inverse gamma distribution (Zellner 1971) with shape parameter $\alpha/2$ and scale parameter $\psi/2$. That is, the error precisions, which are the inverse of the error variances, have a gamma distribution with mean α/ψ and variance $2\alpha/\psi^2$. We use standard prior distributions, which are given in Appendix A. This appendix also describes the estimation procedure.

A posterior analysis reveals that the design matrices X_i in Equation (1) can be less than full rank. The HB estimator of the partworths is a convex function of an individual-level estimator and a pooled estimator where the weights depend on the accuracy of these estimators. Using standard Bayesian arguments gives the posterior mean of subject i's partworth:

$$E(\beta_i \mid Y) = E_{\{\sigma_i\},\Lambda \mid Y}[D_i(\sigma_i^{-2} X_i' Y_i + \Lambda^{-1} U_n z_i)], \tag{3}$$

where $D_i = (\sigma_i^{-2} X_i' X_i + \Lambda^{-1})^{-1}$ is the posterior variance of β_i; U_n is the posterior mean of Θ:

$$\mathrm{vec}(U_n) = V_n\left[\sum_{i=1}^{n}\left(z_i \otimes X_i' \sum_i^{-1} Y_i\right) + V_0^{-1}\,\mathrm{vec}(U_0)\right];$$

$\mathrm{vec}(U)$ stacks the columns of U into a vector; "$\otimes$" is the Kronecker product; U_0 is the prior mean of Θ; V_0 is the prior variance of Θ;

$$V_n = \left(\sum_{i=1}^{n}\left[z_i z_i' \otimes X_i' \sum_i^{-1} X_i\right] + V_0^{-1}\right)^{-1}$$

is the posterior variance of Θ; and $\Sigma_i = \sigma_i^2 I + X_i \Lambda X_i'$ is the variance of Y_i after integrating over β_i.

Because U_n is the posterior mean of Θ based on all of the subjects, $U_n z_i$ is a pooled estimator of β_i. Assuming that the design matrices have full rank, Equation (3) can be rewritten:

$$E(\beta_i \mid Y) = E_{\{\sigma_i\},\Lambda \mid Y}[W\hat{\beta}_i + (I - W)U_n z_i)],$$

where $W = (\sigma_i^{-2}X_i'X_i + \Lambda^{-1})^{-1}\sigma_i^{-2}X_i'X_i$, and $\hat{\beta}_i$ is the ordinary least square (OLS) estimator of β_i based on the data from subject i. The tradeoff between the OLS and pooled estimators is determined by the weights W. If each subject responds to a large number of profiles, W is approximately the identity matrix, and the HB estimator is essentially equivalent to the OLS estimator. If there are few profiles per subject, the OLS is inaccurate, and the HB estimator relies more heavily on the pooled estimator, $U_n z_i$. HB estimators are more efficient, in terms of mean squared error, than OLS estimators (cf. Judge et al. 1985, Chapter 3).

In HB estimation the design matrices do not have to have full rank, unlike OLS applied at the individual level. In particular, Equation (3) can be computed for design matrices with less than full rank because $D_i = (\sigma_i^{-2}X_i'X_i + \Lambda^{-1})^{-1}$ exists even if $X_i'X_i$ is singular. This fact leads to the possibility of giving each subject fewer profiles to evaluate than the number of partworths. The hierarchical Bayes analysis creates the opportunity to recover both the heterogeneity in partworths and individual-level partworths, even when the number of responses per subject is less than the number of parameters per subject. The following sections empirically and analytically explore this property.

3. Empirical Examples

This section presents an empirical investigation of the recovery of partworth heterogeneity as the number of observations per subject are decreased. The first example is a simulation study using synthetic data, and the second is a study of personal computer design. After obtaining responses from the full design, both studies randomly delete profiles. The random deletion provides an acid test for the recovery of heterogeneity because it does not take advantage of a particular experimental design or individual-level features in allocating profiles to subjects. The personal computer survey does not manipulate the number of profiles presented to the subjects and, thus, cannot be used to determine possible biases due the length of the survey.

3.1. Synthetic Data Example

This section uses synthetic data generated from a known model to illustrate the recovery of partworth heterogeneity. The intent of this study is not to demonstrate that Bayes estimators are superior to ordinary least squares (OLS) estimators because that issue has been conclusively settled elsewhere (cf. Judge et al. 1985, Chapter 3). The intent is to demonstrate that HB methods can be used when OLS fails due to insufficient data. The synthetic data were constructed so that the individual-level OLS estimators of the partworths would be accurate when using the full design matrix.

In the full design, simulated responses from 100 subjects with 16 profiles per subject were generated from a common design matrix. The experiment employed five binary attributes for a 2^{5-1} fractional factorial design in which the last factor is confounded with the four-way interaction, assumed to be zero, of the first four factors. The effects-coded design matrix accommodates main effects for the five factors. The OLS estimates of six parameters using 16 observations in an orthogonal design are accurate. The study also uses a validation sample in which each subject responds to four profiles. The validation design was selected to generate unequal market shares.

Individual-level partworths and error variances were independently generated from normal and inverse gamma distributions, respectively, for each subject. The means and variances that were used to generate the individual-level parameters are: $\Theta = (10, 1, 2, 3, -1, 0)'$; $\text{diag}(\Lambda) = (1, 4, 4, 16, 25, 1)\lambda_{i,j} = 2.8$ if $i = 2, j = 3$ or $i = 3, j = 2$ and 0 otherwise; $E(\sigma^2) = 5$ and $\text{var}(\sigma^2) = 9$. In Equation (2), the covariate z_i is equal to one, and Θ is the population mean of the individual-level partworths. After generating the individual level partworths and error variances, the calibration and validation data were generated from Equation (1) using the common design matrix and validation profiles.

The full dataset with 16 profiles per subject was first analyzed by ordinary least squares (OLS) for each subject and then by hierarchical Bayes (HB) methods. Next, two observations per subject were randomly deleted. The HB model was refitted using the new design matrices. This procedure was repeated, so that the number of profiles per subject decreased from 16 to two in increments of two. Because observations were randomly deleted, some of the subject-level design matrices were less than full rank with 10 observations per subject. Up to this point, the OLS and HB estimates were fairly similar.

Table 1 presents performance measures for the calibration sample in the top half and for the validation sample in the bottom half. Columns two to five for the calibration sample gives the root-mean-squared error (RMSE) and correlation between the true, individual level parameters and their estimates, and the sixth and seventh columns list the RMSE between the true and estimated partworths' means, Θ, and the nonzero elements of the covariance matrix, Λ. (We excluded the zero elements because their prior means are zero.) For OLS these heterogeneity parameters were estimated by the mean and covariance of the individual level estimates. With 16 profiles per subject, there is little to distinguish between the OLS and HB estimators. As the number of profiles decreases, the individual-level estimates become less accurate, as one would expect. Comparing the accuracy of the individual-level estimates (column 2) with that of heterogeneity parameters, Θ and Λ (last two columns), indicates that HB is substantially more accurate in estimating the heterogeneity parameters. The RMSEs are lower and increase more gradually than those for the individual-level partworths.

The lower half of Table 1 presents out-of-sample performance measures based on the validation sample. The RMSE and the correlation of the actual and predicted responses are equivalent for the OLS and HB estimates using the full dataset. The correlation is above 0.8 until there are two profiles per subject, and the average change in RMSE is 6% until four profiles and 9% until two profiles. The fourth column reports the "hit rate," which is the proportion of times that the predicted maximum utility profile corresponds to the actual maximum in the validation sample. The hit rate for the HB is 93% compared to 90% for the OLS, and it tends to decrease as profiles are deleted, reaching a low of 60% with only two profiles per subject. The RMSE between the observed and predicted market shares is given in the last column. OLS and HB have similar performance with 16 profiles, and HB actually is better with eight to 14 profiles, which is an anomaly due to randomly deleting calibration profiles.

Two points from the simulation should be emphasized. First, HB provides both individual-level estimates and partworth heterogeneity estimates even when OLS cannot. Second, the heterogeneity estimators are more accurate than the individual ones. These observations imply that properly designed, short questionnaires can

Table 1 Calibration and Validation Performance Measures for the Simulation Study

Calibration Sample

	Partworths		Error Variance		RMSE of Partworths	
Profiles	RMSE	Corr†	RMSE	Corr	Mean	Cov
Ordinary Least Squares						
16	0.581	0.945	2.980	0.742	0.175	1.540
Hierarchical Bayes						
16	0.540	0.946	2.790	0.704	0.178	1.442
14	0.580	0.940	3.136	0.661	0.179	1.388
12	0.621	0.931	3.612	0.579	0.189	1.411
10	0.723	0.912	4.346	0.549	0.174	1.380
8	0.866	0.876	5.080	0.506	0.190	1.737
6	1.074	0.826	6.615	0.365	0.205	1.803
4	1.591	0.710	6.954	0.127	0.280	2.578
2	2.435	0.484	4.973	0.167	0.410	6.444

Validation Sample

	Prediction			
Profiles	RMSE	Corr	Hit Rates‡	Market Shares RMSE
Ordinary Least Squares				
16	2.603	0.939	0.90	0.051
Hierarchical Bayes				
16	2.605	0.938	0.93	0.051
14	2.676	0.934	0.95	0.024
12	2.694	0.933	0.95	0.045
10	2.846	0.925	0.91	0.024
8	3.181	0.906	0.86	0.047
6	3.465	0.887	0.84	0.079
4	4.305	0.821	0.68	0.065
2	6.271	0.593	0.60	0.158

† Average correlation across factors.

‡ Proportion of subjects where the maximum predicted validation response corresponded to the actual maximum.

be used to obtain heterogeneity estimates, even though the individual-level estimates are not accurate. In the simulation, we randomly deleted profiles, but we do not recommend this as a design strategy. Based on the analysis in §4, we recommend using blocked factorial designs.

3.2. Conjoint Analysis of Personal Computers

This section describes a conjoint study that utilizes individual-level covariates to describe the heterogeneity in the partworths. The subjects were first-year students in the Masters of Business Administration program at the University of Michigan. The data were collected during the 1994 fall semester. The subjects rated the likelihood of purchasing hypothetical personal computers on an 11-point scale (0 to 10) and provided information about their experience with computers, self-assessments of their expertise, and demographic information. A lottery system with cash prizes encouraged participation. Of a class of approximately 425 students, 201 responded. After eliminating subjects with missing responses, 179 subjects are used in the following analysis. By industrial standards, our conjoint study had a small number of profiles—only 20. Yet the response rate was less than 50%, and 10.9% of those who responded did not complete the questionnaire.

Table 2 presents the 13 factors used to describe the personal computers and the six individual-level covariates. The experiment is an orthogonal, effects-coded design with main effects only and 16 calibration profiles and four validation profiles, which were selected to mimic computer systems advertised in major trade magazines. The means and standard deviations for the covariates are also reported.

The computer descriptions involved both intrinsic (technical) features, such as the amount of RAM and CPU speed, and extrinsic features, such as telephone support and distribution channel. As purported in DeSarbo et al. (1990), our presupposition was that students with higher computer expertise and technical backgrounds would find the intrinsic factors more important than those with less computer expertise and nontechnical backgrounds.

Table 3 presents the posterior means and standard deviations of the regression coefficients Θ that relate the individual-level covariates to the partworths in the HB model. Many of the coefficients have posterior means that are one or more posterior standard deviations from zero. Although none of the covariates are definitive

Table 2 MBA Computer Conjoint Analysis

A. Telephone Service Hot Line	H. Color of Unit
−1 = No	−1 = Beige
1 = Yes	1 = Black
B. Amount of RAM	I. Availability
−1 = 8 MB	−1 = Mail order only
1 = 16 MB	1 = Computer store only
C. Screen Size	J. Warranty
−1 = 14 inch	−1 = 1 year
1 = 17 inch	1 = 3 year
D. CPU Speed	K. Bundled Productivity Software
−1 = 50 MHz	−1 = No
1 = 100 MHz	1 = Yes
E. Hard Disk Size	L. Money Back Guarantee
−1 = 340 MB	−1 = None
1 = 730 MB	1 = Up to 30 days
F. CD ROM/Multimedia	M. Price
−1 = No	−1 = $2000
1 = Yes	1 = $3500
G. Cache	
−1 = 128 KB	
−1 = 256 KB	

Subject Level Covariates

Variable	Description	Mean	STD
FEMALE	0 if male and 1 if female	0.27	0.45
YEARS	Years of full-time work experience	4.4	2.4
OWN	1 if own or lease a microcomputer and 0 otherwise	0.88	0.33
TECH	1 if engineer or computer professional 0 otherwise	0.27	0.45
APPLY	Number software applications	4.3	1.6
EXPERT	Sum of two self-evaluations. Each evaluation in on a five-point scale with 1 = Strongly Disagree, 3 = Neutral, and 5 = Strongly Agree. The first evaluation is, "When it comes to *purchasing* a microcomputer, I consider myself pretty knowledgeable about the microcomputer market." The second is, "when it comes to *using* a microcomputer, I consider myself pretty knowledgeable about microcomputers."	7.6	1.9

Number of subjects:	179
Number of calibration profiles per subject:	16
Number of validation profiles per subject:	4

Table 3 Sensitivity of Partworths to Subject Level Covariates Using 16 Profiles per Subject (posterior standard deviations are in parentheses)

	Covariate						
Variable	Intercept	FEMALE	YEARS	OWN	TECH	APPLY	EXPERT
Intercept	3.698**	−0.043	−0.111**	−0.158	−0.248	0.112*	0.167**
	(0.598)	(0.271)	(0.049)	(0.347)	(0.271)	(0.080)	(0.071)
A Hot Line	−0.047	0.226**	−0.002	−0.105	−0.019	−0.004	0.026*
	(0.195)	(0.087)	(0.016)	(0.115)	(0.084)	(0.025)	(0.023)
B RAM	0.515**	−0.085	−0.003	0.139*	0.168*	0.043*	−0.065**
	(0.208)	(0.093)	(0.017)	(0.127)	(0.086)	(0.027)	(0.024)
C Screen Size	0.058	−0.055	−0.009	0.044	0.109*	0.005	0.013
	(0.176)	(0.079)	(0.014)	(0.102)	(0.078)	(0.022)	(0.020)
D CPU	−0.167	−0.101	−0.026*	0.158	0.171*	0.014	0.059*
	(0.279)	(0.131)	(0.023)	(0.172)	(0.127)	(0.038)	(0.033)
E Hard Disk	0.013	−0.157*	−0.014	0.037	0.060	0.017	0.015
	(0.183)	(0.082)	(0.014)	(0.105)	(0.080)	(0.023)	(0.021)
F CD ROM	0.591**	−0.164*	−0.010	−0.062	−0.075	0.015	0.001
	(0.251)	(0.113)	(0.020)	(0.148)	(0.107)	(0.033)	(0.029)
G Cache	−0.266*	−0.043	−0.004	0.127*	0.019	−0.036*	0.049**
	(0.192)	(0.092)	(0.015)	(0.118)	(0.087)	(0.026)	(0.023)
H Color	0.274*	−0.047	−0.004	0.017	−0.095*	−0.014	−0.019*
	(0.160)	(0.070)	(0.013)	(0.093)	(0.072)	(0.021)	(0.019)
I Availability	0.157*	0.037	0.021*	0.138*	−0.097*	−0.011	−0.029*
	(0.156)	(0.068)	(0.013)	(0.092)	(0.070)	(0.021)	(0.018)
J Warranty	−0.089	0.149*	0.024*	0.029	0.008	0.026*	−0.010
	(0.167)	(0.079)	(0.015)	(0.100)	(0.072)	(0.022)	(0.020)
K Software	0.315*	0.009	−0.032**	−0.034	0.101*	0.010	−0.004
	(0.179)	(0.081)	(0.014)	(0.104)	(0.079)	(0.023)	(0.020)
L Guarantee	0.023	0.031	0.025*	−0.117*	−0.081	0.013	0.004
	(0.185)	(0.085)	(0.015)	(0.107)	(0.081)	(0.025)	(0.022)
M Price	−1.560**	0.385**	0.040*	−0.176	−0.064	0.001	0.041
	(0.398)	(0.173)	(0.031)	(0.233)	(0.170)	(0.052)	(0.047)

* The posterior mean is at least one posterior standard deviation from zero.

** The posterior mean is at least two posterior standard deviations from zero.

drivers of the variation in the partworths, at least two broad themes emerge. The first is that technically knowledgeable users tend to value intrinsic features, such as CPU speed, more than less knowledgeable users. The second theme, which we did not anticipate, is that students with more work experience value extrinsic features more than those with less work experience. One is tempted to conjecture that more experienced students have a broader view of purchasing a computer and believe that extrinsic support services are important components of the productive use of their capital investment.

As a basis of comparison, we fitted a latent class metric (LCM) conjoint model (DeSarbo et al. 1992) to the full data. LCM is a competing semi-parametric methodology for full profile conjoint analysis. It represents heterogeneity with discrete support points (market segments) where partworths are estimated by latent classes, and each individual belongs to each class with some probability. Four latent classes are optimal using the CAIC criterion.

Next, we investigated the HB and LCM procedures' ability to recovery partworth heterogeneity as profiles were randomly deleted from each subject. Table 4 re-

Table 4 Calibration and Validation Sample Performance Measures for the Computer Survey

Calibration Sample			Validation Sample			
	RMSE† for Partworth		Prediction			Market
Profiles	Means	STD	RMSE	Corr	Hit Rates‡	Shares RMSE
Ordinary Least Squares			Ordinary Least Squares			
16	0.010	0.180	1.998	0.7152	0.637	0.088
Hierarchical Bayes			Hierarchical Bayes			
16	0.000	0.000	1.811	0.7530	0.670	0.069
12	0.020	0.041	1.851	0.7425	0.687	0.123
8	0.045	0.064	1.983	0.7029	0.654	0.106
4	0.066	0.137	2.262	0.5877	0.587	0.223
Latent Class			Latent Class			
16	0.009	0.221	4.048	0.6834	0.380	0.448
12	0.027	0.210	4.194	0.6496	0.408	0.330
8	0.046	0.158	4.470	0.5838	0.291	0.268
4	0.094	0.283	4.928	0.4800	0.374	0.148

† Root mean squared error with respect to the hierarchical Bayes estimate using 16 profiles.

‡ Proportion of subjects where the maximum predicted validation response corresponded to the actual maximum.

ports the performance of OLS, HB, and LCM in the calibration and the validation samples. The true partworth means and standard deviations are not known, so the RMSEs were computed with respect to the HB estimates with 16 profiles per subject. The OLS, HB, and LCM estimates of the partworths' means are equivalent when using 16 profiles, while there are systematic differences in the standard deviations of the partworths in the sample. The HB estimates of the partworths' standard deviations tend to be slightly less than the OLS. This result is expected because one motivation for using HB or other shrinkage estimators is to decrease the sampling variation when estimating many parameters. Thus, OLS tends to overstate the heterogeneity in the partworths because of larger sampling variation. LCM greatly understates the heterogeneity in the partworths as compared to OLS and HB because LCM does not admit individual partworths outside of the convex hull of the support points.

We then refitted the model using 12, 8, and 4 profiles per subject where profiles were randomly deleted. Individual level OLS estimates do not exist for the these sample sizes. The LCM estimates of the standard deviations continue to be much less than the HB estimates.

The second part of Table 4 compares the predictive performance of the OLS, HB, and LCM estimators with the validation sample. The HB predictions of the subjects' responses to the validation profiles using 16 and 12 calibration profiles have larger correlations and smaller RMSEs than the OLS predictions using 16 profiles. In fact, there is no noticeable decline in the predictive accuracy of HB until there are only 4 calibration profiles per subject, at which point the RMSE increases by less than 25% despite having 75% fewer observations. The HB hit rate with eight or more profiles exceeds the OLS hit rate using 16. This result may be due to HB incorporating the information from the multivariate regression of the partworths on the covariates in Equation (2). The RMSE between the actual and predicted market shares is better for HB than OLS when using the entire dataset. Not surprisingly, the RMSE increases as profiles are deleted. HB with 12 profiles is off by nearly twice the percentage points as HB with 16 profiles.

If the information overload hypothesis were true, it may be the case that the predictive accuracy actually increases with a decrease in the number of profiles per subject in a study that varied the number of profiles. This study remains silent on the impact of questionnaire length on response quality because a subject's responses were randomly deleted after he or she responded to the entire questionnaire, and the number of profiles were not physically manipulated during the experiment.

The LCM predictions perform substantially worse than HB across all measures. As previously noted, the estimated individual-level partworths in LCM must belong to the convex hull of the support points for the market segments. Thus, LCM does reasonably well in estimating the mean partworth but tends to underestimate their dispersion. Consequently, predictions using LCM fail for subjects' whose true preferences are outside the convex hull.

In conclusion, hierarchical Bayes models can recover heterogeneity even when there are not sufficient observations to obtain subject level estimates by more standard methods. The next section generalizes these empirical observations with an analytical analysis of the tradeoff between the number of subjects and the number of profiles per subject.

4. Subject-profile Tradeoff

Section 3 provided some empirical evidence that HB conjoint analysis is capable of estimating the heterogeneity in the partworths when each subject receives relatively few profiles in relation to the number of attributes. This section generalizes these observations by investigating the effect of varying the number of subjects and the number of profiles per subject on the Fisher's information of the design. The asymptotic variance of the estimated heterogeneity is the inverse of the Fisher's information. We compute the Fisher's information for the general case, then we find the D-optimal design for two cases: orthogonal designs and blocked factorial designs. A design, within a given class, is D-optimal if it maximizes the determinant of the Fisher's information subject to a cost constraint. Pilz (1991) considers the Bayesian design of experiments for the fixed effects, linear model; this section considers random effects models.

In the context of the random effects model specified by Equations (1) and (2), we focus on the information matrix for Θ and Λ, the parameters that describe the heterogeneity of the partworths. The Fisher's information matrix has cells:

$$FI(\gamma_1, \gamma_2) = -E\left[\frac{\partial^2}{\partial\gamma_1\partial\gamma_2}\log f(Y|\Theta, \Lambda)\right],$$

where γ_1 and γ_2 are any of the parameters in Θ and Λ; the expectation is with respect to Y given Θ and Λ; and $f(Y|\Theta, \Lambda)$ is the marginal density of Y given Θ and Λ. This information matrix implicitly depends on the number of subjects and the number of profiles per subject. Thus, the tradeoff between the number of subjects and profiles can be evaluated for different conjoint studies.

THEOREM 1. *Write the design matrices, $\{X_i\}$, as column vectors: $X_i = [x_{i,1}, \ldots, x_{i,p}]$, and define $\Sigma_i = \sigma_i^2 I_{m_i} + X_i\Lambda X_i'$. The Fisher's information matrix has the following entries:*

$$FI(\Theta, \Theta) = \sum_{i=1}^{n} (z_i z_i' \otimes X_i'\Sigma_i^{-1}X_i),$$

$$FI(\Theta, \Lambda) = 0 = FI(\Lambda, \Theta)',$$

$$FI(\lambda_{u,u}, \lambda_{r,r}) = \tfrac{1}{2}\sum_{i=1}^{n} (x_{i,u}'\Sigma_i^{-1}x_{i,r})^2,$$

$$FI(\lambda_{u,u}, \lambda_{r,s}) = \sum_{i=1}^{n} (x_{i,u}'\Sigma_i^{-1}x_{i,r})(x_{i,u}'\Sigma_i^{-1}x_{i,s}) \quad \text{if } r \neq s,$$

$$FI(\lambda_{u,v}, \lambda_{r,s}) = \sum_{i=1}^{n} (x_{i,u}'\Sigma_i^{-1}x_{i,r})(x_{i,v}'\Sigma_i^{-1}x_{i,s}) + (x_{i,u}'\Sigma_i^{-1}x_{i,s})(x_{i,v}'\Sigma_i^{-1}x_{i,r}) \quad \text{if } u \neq v;\ r \neq s.$$

The proof is in Appendix B. The Fisher's information depends on unknown variances, which can be provided by expert opinion, from past data, or through a pilot experiment.

The next corollary computes the Fisher's information when all subjects respond to a common orthogonal design.

COROLLARY 1. *Assume: (i) all subjects receive a common, orthogonal design matrix X where $X'X = mI$; (ii) the error variances for all subjects is σ_0^2, and (iii) $\Lambda = \lambda_0^2 I$. Define Z to be the n by q matrix of covariates with $Z' = [z_1, \ldots, z_n]$. Then, the determinant of the Fisher's information is*

$$\log|FI| = p\log|n^{-1}Z'Z)| + pq\log\left(\frac{nm}{\sigma_0^2 + \lambda_0^2 m}\right) - p\log(2) + \frac{p(p+1)}{2}\log\left[n\left(\frac{m}{\sigma_0^2 + \lambda_0^2 m}\right)^2\right]. \quad (4)$$

Equation (4) separates the covariates Z from the experimental design X, which greatly simplifies the analysis. It is well known from D-optimal designs for linear models that the first term, $\log(n^{-1}Z'Z)$, is maximized by setting the covariates at their extreme values, although in practice experimenters may not have the luxury of setting all of the covariates and, even if they can, may decide to sample intermediate values to detect nonlinearities or appropriate transformations of the covariates. Because the elements of $n^{-1}Z'Z$ are averages of the covariates over subjects, it does not play an important role in determining the number of subjects. Also, Z does not depend on the number of profiles per subject.

The next theorem computes the D-optimal design when all subjects respond to a common orthogonal design subject to the budget constraint:

$$cn + dmn \leq c_T, \quad (5)$$

where c_T is the total budgeted cost, c is cost for one subject, and d is the cost for one profile per subject; n is the total number of subjects, and m is the number of profiles per subject. There does not exist orthogonal design for every integer. Recall that the dimension of the partworths is p, so $m \geq p$. Define L to be the smallest m that admits an orthogonal design. Similarly, there has to be at least one subject, which implies an upper bound on m due to the budget constraint. Let U be the largest m that admits an orthogonal design and meets the budget constraint.

THEOREM 2. *Assume that the conditions of Corollary* 1 *hold and that the cost constraint is given in Equation* (5). *Define*

$$m^* = \frac{p + 1 + [(p + 1)^2 + 8(p + q + 1) \times (p + 2q + 1)(c/d)(\lambda_0^2/\sigma_0^2)]^{1/2}}{2(p + 2q + 1)(\lambda_0^2/\sigma_0^2)},$$

$$a = \frac{2(p + q + 1)c/d + U(p + 1)}{U^2(p + 2q + 1)},$$

$$b = \frac{2(p + q + 1)c/d + L(p + 1)}{L^2(p + 2q + 1)}. \qquad (6)$$

Given Z, the optimal number of profiles per subject, $\tilde{m}$, is the lower bound, L, if $\lambda_0^2/\sigma_0^2 \geq b$; the upper bound, U, if $\lambda_0^2/\sigma_0^2 \leq a$; or one of the two orthogonal designs with m closest to m^* if $a < \lambda_0^2/\sigma_0^2 < b$. The optimal number of subjects is the greatest integer that satisfies the budget constraint given the optimal number of profiles. That is, the optimal n is the greatest integer less than or equal to $c_T/(c + d\tilde{m})$.

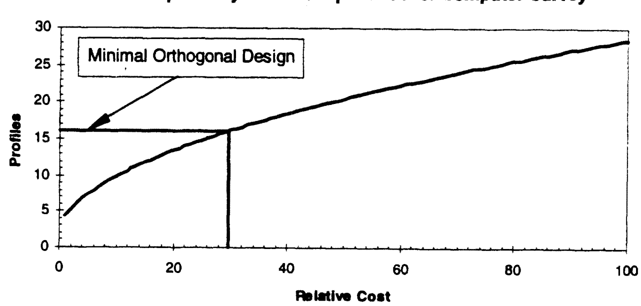
The optimal number of profiles per subject in Theorem 2 has an intuitive interpretation. If the unexplained heterogeneity λ_0^2 relative to the error variance σ_0^2 exceeds a linear function of the cost ratio c/d, then the optimal design uses the minimum m and the maximum n to obtain the best estimate of the heterogeneity, even though the individual-level partworths may not be estimated with great precision. If the unexplained heterogeneity is small relative to the error variance, then the optimal design uses the maximum m and the minimum n. Here, accurately estimating the individual-level partworths for only a few subjects leads to accurate estimators of their heterogeneity.

For the computer survey, the average error variance across subjects is 1.668, and the average of the $\lambda'_{u,u}$s is 0.3431 for a λ_0^2/σ_0^2 of 0.2057. Figure 1 graphs m^* in Equation (6) as a function of the relative costs c/d. If the relative cost is less than 29.25, then the optimal number of profiles per subject is 16, the minimal orthogonal design.

The previous analysis assumed that all subjects receive a common, orthogonal design matrix where the number of profiles per subject is greater than the number of regression parameters. Next, we will investigate a more complex example of incomplete designs where subjects receive fewer profiles than regression parameters. In this case, different sets of profiles have to be administered to different groups of subjects to acquire information about the heterogeneity in all of the partworths. For the important special case of blocked factorial designs, explicit expressions of the Fisher's information as functions of the number of subjects and the number of observations per subject are presented in Theorem 3. The main conclusion of this analysis is that it is possible to estimate accurately the partworth heterogeneity in situations where OLS at the subject level do not exist.

A blocked design is a subset of a full factorial design where the profiles for the design are selected by one or more blocking variables. f binary attributes lead to 2^f profiles for the full design X. The profiles in the full design are divided into 2^b blocks with 2^{f-b} profiles per block where b factors are used as blocking variables. Let X_i, for $i = 1$ to 2^b, be the design matrices for the blocks with their columns arranged by the alias classes generated by the blocking scheme. Then the inner-product of X_i is a block diagonal matrix:

Figure 1 **Optimal Number of Items per Subject as a Function of Relative Cost per Subject to Cost per Item for Computer Survey**



$$X_i' X_i = 2^{f-b} \operatorname{block}(P_i, \ldots, P_i),$$

and P_i is a $2^b \times 2^b$ matrix of positive and negative ones (John 1971, p. 158–165).

Next, we compute the Fisher's information for the class $\mathcal{BF}$ that has the following structure. Each subject belongs to one of q classes, such as gender and age group. The individual-level covariates,

$$z_i = (1, G_{i,1}, \ldots, G_{i,q-1})',$$

denote subject i's class membership based on the dummy variables: $G_{i,k} = 1$ if subject i belongs to class k, and 0 otherwise. The design has 2^f parameters. It could have f binary attributes and estimate all of the main effects and iteractions, or it could have more than f attributes and identify some of the main effects with higher order interactions. Each of the 2^{b_k} blocks of a 2^{f-b_k} factorial experiment are allocated n_k subjects in class k for $k = 1$ to q.

THEOREM 3. *Consider the class $\mathcal{BF}$ of blocked factorial designs described above. Assume that the error variances for all subjects within class k is $\sigma_{0,k}^2$ and that $\Lambda = \lambda_0^2 I$. Then the Fisher's information for Θ and its determinant follow*:

$$FI(\Theta, \Theta) = \begin{bmatrix} \sum_{k=1}^{q} n_k a_k & n_1 a_1 & n_2 a_2 & \cdots & n_{q-1} a_{q-1} \\ n_1 a_1 & n_1 a_1 & 0 & \cdots & 0 \\ n_2 a_2 & 0 & n_2 a_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_{q-1} a_{q-1} & 0 & 0 & \cdots & n_{q-1} a_{q-1} \end{bmatrix} \otimes I_p,$$

$$\det[FI(\Theta, \Theta)] = \left(\prod_{k=1}^{q} n_k a_k \right)^p,$$

where $a_k = (2^{-f}\sigma_{0,k}^2 + \lambda_0^2)^{-1}$, and $p = 2^f$. The Fisher's information for $\lambda_{u,u}$ is

$$FI(\lambda_{u,u}, \lambda_{u,u}) = \frac{1}{2} \sum_{k=1}^{q} n_k 2^{-b_k} a_k^2.$$

The proof is in Appendix B.

Theorem 3 can be used to compare two experiments when $q = 1$: (1) n_F subjects each evaluate the full 2^f design, and (2) $n_B 2^b$ subjects evaluate the blocked 2^{f-b} designs where each of the 2^b blocks is administered to n_B subjects. The Fisher's information matrices for Θ for the two experiments are equal if $n_B = n_F$ and for the partworths' variances if $n_B = n_F 2^b$. In other words, if there are a total of $n2^f$ profiles to be allocated among subjects, then the estimation accuracy for the means of the partworths remains constant whether n subjects each respond to 2^f profiles from the full design or $n2^b$ subjects respond to 2^{f-b} profiles from the blocked designs with n subjects per block. However, the second experiment would have less accuracy in estimating the partworths' variances.

A slightly different analysis is to consider values of n and b that lead to the same value of the information measure:

$$g(n, b) = v \sum_{k=1}^{q} \log(n_k a_k) + w \log\left(\frac{1}{2} \sum_{k=1}^{q} n_k 2^{-b_k} a_k^2 \right), \quad (7)$$

where v and w are nonnegative constants that balance the relative importance of estimating Θ and the unexplained variances $\{\lambda_{u,u}\}$ of the partworths. When $q = 1$, setting $g(n, b) = c$, and solving for n results in

$$n = 2^{(b+1)w/(v+w)} (2^{-f}\sigma_0^2 + \lambda_0^2)^{(v+2w)/(v+w)} \exp[c/(v + w)].$$

This equation gives n, the number of subjects per block, to maintain a given information content for different numbers of blocks. As the number of blocks 2^b increases and the number of profiles 2^{f-b} per subject decreases, the number of subjects per block increases. The main conclusion is that by carefully deploying blocked factorial design, the marketing researcher can recover partworth heterogeneity while using fewer observations per subject than regression parameters.

One method to determine the design requires the specification of the widths of the approximate 95% highest posterior Bayesian (HPB) intervals (cf. Berger 1985) for the mean partworth and the unexplained heterogeneity. HPB intervals are the Bayesian counterpart to confidence intervals. Using the normal approximation to the posterior distribution, they have the form: the posterior mean plus and minus 2 times the square root of the inverse Fishers' information. Consider the simple case when q is one in Theorem 3. If the total width of the 95% HPB interval for θ_k is $2v$, then the minimum number of subjects per block n to meet this requirement is the smallest integer greater than or equal to $4(2^{-f}\sigma_0^2 + \lambda_0^2)/v^2$. Figure 2a graphs the number of subjects versus the half-width v for the computer survey. Similarly, the number of blocks 2^b can be determined by specifying the total width, $2w$ for the approximate 95% interval for $\lambda_{k,k}$. Then b is the minimum of $f - 1$ and the smallest nonnegative integer greater than

$$\ln(0.5(2^{-f}\sigma_0^2 + \lambda_0^2)^{-1}w^2/v^2)/\ln(2).$$

Figure 2b graphs b as a function of the relative widths w/v for the computer survey. In this example, if the required width for the variance is one-half that of the mean, then the full design should be used, while if it is twice that of the mean, the optimal number of blocks is 8.

The D-optimal design maximizes the determinant, or a function thereof, of the Fisher's information for the heterogeneity parameters Θ and Λ subject to a cost constraint:

$$c_T = \sum_{k=1}^{q} n_k 2^{b_k}[c_k + d_k(2^{f-b_k})], \tag{8}$$

where c_T is the total cost; c_k is the cost of obtaining one subject from group k, and the function $d_k(x)$ is the cost of x profiles per subject in group k. The next theorem provides a procedure for the optimal design.

THEOREM 4. *Suppose that the assumptions of Theorem 3 hold. For nonnegative constants v and w the D-optimal conjoint study that maximizes Equation (7) among the class $\mathcal{BF}$ subject to the cost constraint in Equation (8) can be determined by the following procedure:* (1) *Fix values of $b_1, \ldots, b_q$.* (2) *The optimal n_k's satisfy the following system of quadratic equations for $j = 1$ to q:*

$$v\sum_{k=1}^{q} n_k 2^{-b_k}a_k^2 + wn_j 2^{-b_j}a_j^2$$
$$-\frac{vq + w}{c_T}\left(\sum_{k=1}^{q} n_k 2^{-b_k}a_k^2\right)n_j 2^{b_j}[c_j + d_j(2^{f-b_j})] = 0.$$

(3) *Repeat steps* (1) *and* (2) *for all values of $b_1, \ldots, b_q$, and select the design that optimizes g.*

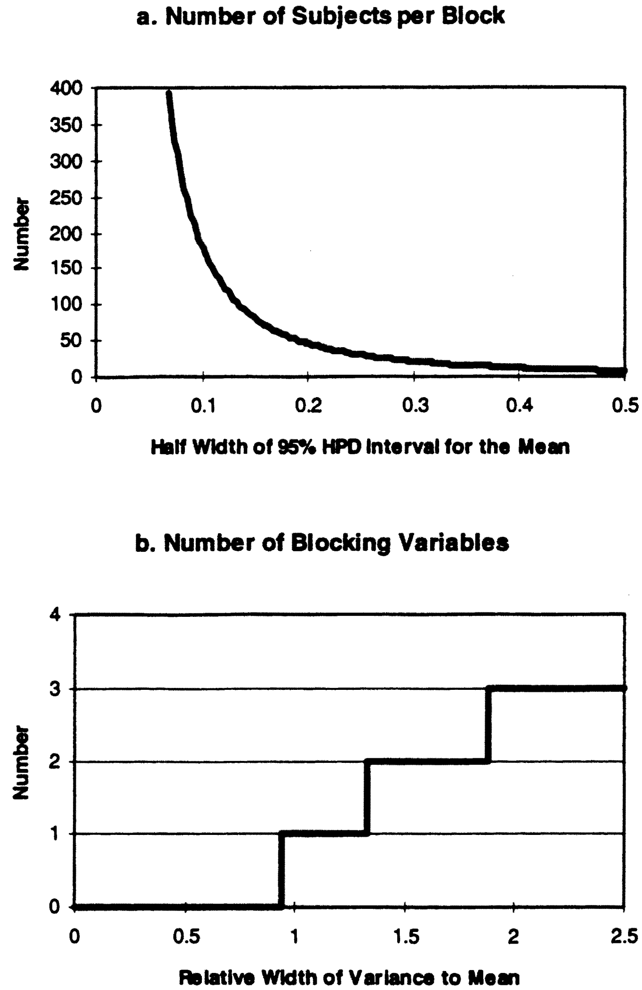
The proof is in Appendix B. The procedure requires an exhaustive search over the blocking variables, which is not too punitive because they can take at most fq values.

The analysis in this section does not include the possibility that response quality degrades as the questionnaire length increases and the complexity of the product or service stimuli increases. One possible approach that includes response bias in the analysis uses cost functions that are convex and increasing in the number of profiles per subject to reflect the additional motivation that subjects would require to carefully evaluate all of the profiles. This cost may be additional monetary compensation to the subjects or may reflect additional expense to provide stimuli, such as actual mock-up of the products or an initial training phase. A different approach would be to model the error variance or the

Figure 2 **Number of Subjects per Block and Number of Blocks for Computer Survey:**

(a) number of subjects per block vs. width of 95% HPB interval for mean partworth

(b) number of blocks vs. ratio of widths of 95% HPB interval for unexplained heterogeneity and 95% HPB interval of mean partworth



unexplained heterogeneity variance as increasing functions of the number of profiles per questionnaire or the number of dimensions to be evaluated. This route, though not without appeal, would first require empirical evidence about the sensitivity of these parameters to conjoint design factors.

A different approach, suggested to us by the area editor and an anonymous reviewer, involves conducting a pilot study with the full design and then analyzing the pilot with both the full design and reduced designs. The accuracy of the reduced design could be evaluated relative to the full design by considering criteria based on either the posterior means and standard deviations for Θ or Λ or the predictive accuracy. After selecting the design based on the pilot, the full scale conjoint study could be administered. This procedure would be particularly useful for large studies that use a pilot study. In this case, the study would not incur additional cost due to the pilot, and the full study may have increased efficiency. A possible limitation to this approach is that if the full design has response bias, such as straight-line stereotyping, it will carryover to the reduced design, and the criterion may be misleading.

5. Discussion

Suppose that a marketing researcher is planning a conjoint study of a complex product or service consisting of many salient attributes. Traditional estimation methods, such as ordinary least squares, would require each subject to evaluate a large number of profiles in order to obtain individual-level partworths. However, in most realistic settings the researcher frequently has other considerations, ranging from logistical considerations and budget constraints to possible biases induced by long questionnaires, as documented in the Introduction, in addition to estimation accuracy. Thus, he or she needs to balance the competing goals of obtaining individual-level partworth estimates with designing an efficient and parsimonious experiment that allows for the estimation of partworth heterogeneity.

This paper provides a framework to design and analyze conjoint studies by employing a random effects model and hierarchical Bayes (HB) methods. It investigates the tradeoff in the estimation accuracy between the number of subjects and the number of profiles per subject and demonstrates that partworth heterogeneity can be recovered when the subjects respond to reduced sets of profiles, even when the number of profiles per subject is less than the number of attributes. The results of this paper enhances the marketing researcher's ability to obtain information about complex products or services.

The results of this paper would be very useful if the information overload hypothesis were true. If it were false, then given sufficient resources, the marketing researcher could design a large study and use traditional estimation techniques. On the other hand, if it were true, then money alone would not suffice to overcome potential response biases induced by a long questionnaire. Despite the preponderance of academic research supporting the information overload hypothesis, it remains controversial due to measurement issues surrounding the quantity and quality of information and the quality of a subject's responses. Additionally, the subjects' motivation and innate ability seem to be important factors. Thus, investigations of the hypothesis should not only vary the parameters of the conjoint study but also control for these subject specific factors. The work of Lee and Yates (1992) may provide useful perspectives on these issues. Neither the empirical studies nor the mathematical analysis of this paper addresses or relies on the information overload hypothesis. The paper addresses statistical issues of conjoint design and analysis, not the important, psychological issues about the effects of the conjoint design on the subjects' responses.

In addition to further research about information overload and the psychological affects of conjoint design parameters, there are a number of interesting statistical problems remaining for future research. First, the model in this paper assumes a specific form for the heterogeneity in the individual-level parameters. Further research should investigate alternative specifications. For example, instead of a multivariate regression model, the partworths may follow a mixture model that indicates market segments. The sensitivity of partworth estimation, prediction of validation responses, and market share estimation to the assumptions about the nature of the heterogeneity should be investigated. Second, the generalization of this HB framework to adaptive conjoint analysis, hybrid models, and choice-based conjoint analysis would be desirable given the increasing popularity of these alternative conjoint procedures with the advent of readily accessible commercial software (Johnson 1987 and 1993). Third, a more detailed study concerning the optimal design of conjoint studies for broader classes would be useful. For example, what are the merits of having a "learning" phase, and how long should it be? Finally, more actual conjoint applications need to be performed with comparative validation with other methods (e.g., hybrid models, choice based conjoint, ACA, bridging designs, etc.) which require different data collection protocols to assess the promising benefits of the hierarchical Bayes procedure.

Appendix A

Hierarchical Bayes Estimation

HB inference requires prior distributions on the unknown parameters. The ones that we employ have been widely used in Bayesian analysis (Berger 1985, Box and Tiao 1973, DeGroot 1970, and Zellner 1971). They can flexibly encode prior information about the unknown parameters when substantive prior knowledge exists, and they can be selected so that they do not dominate the posterior distributions when prior knowledge is vague. We assume that the prior for Θ is $N_{p\times q}(U_0, V_0)$, the $p \times q$-dimensional matrix normal distribution with mean U_0 and covariance matrix V_0. U_0 is a $p \times q$ matrix, and V_0 is a $pq \times pq$ positive definite matrix. The matrix normal density is expressed by stacking the columns of Θ and U_0: vec(Θ) is $N_{pq}(\text{vec}(U_0), V_0)$. A noninformative prior for Θ is symbolically equivalent to setting U_0 and V_0^{-1} equal to zero. We use this noninformative prior in the empirical examples of §3. The prior for Λ^{-1} is $W_p(\eta_0, \Delta_0)$, the p-dimensional Wishart distribution (Zellner 1971, p. 389–394) with η_0 degrees of freedom and scale parameter Δ_0. In §3, we set $\eta_0 = p$ and $\Delta_0 = I$. The prior for $\log(\alpha)$ is $N(a_0, d_0^2)$. In §3, we select a_0 and d_0 so that the median of α is 0.1, and the 97th percentile is 100, implying weak prior knowledge about α. Finally, the prior for ψ is $G(r_0/2, s_0/2)$, the gamma distribution with shape parameter $r_0/2$ and scale parameter $s_0/2$. That is, the mean of ψ is r_0/s_0, and its variance is $2r_0/s_0^2$. In §3, r_0 and s_0 are set equal to one.

The posterior distributions do not have closed-form expressions. Fortunately, the model can be analyzed by well-known numerical methods, such as Markov chain Monte Carlo (MCMC), which exploit the special structure of HB models. See Arnold (1993), Gelfand and Smith (1990), Smith and Roberts (1993), Tanner (1993), and Tanner and Wong (1987) for a description of the general procedure. The method iteratively generates random deviates from the posterior distribution of one set of parameters given the current value of all other parameters and the data. The required conditional distributions are given in the following steps.

1. Independently generate β_1 to β_n from normal distributions: $\beta_i \sim N_p(b_i, D_i)$ where $b_i = D_i(\sigma_i^{-2}X_i'Y_i + \Lambda^{-1}\Theta z_i)$, and $D_i = (\sigma_i^{-2}X_i'X_i + \Lambda^{-1})^{-1}$.

2. Generate Θ from $N_{p\times q}(U_n, V_n)$ where $V_n = [ZZ' \otimes \Lambda^{-1} + V_0^{-1}]^{-1}$; $\text{vec}(U_n) = V_n[(Z \otimes \Lambda^{-1})\,\text{vec}(B) + V_0^{-1}\,\text{vec}(U_0)]$; $Z = (z_1, \ldots, z_n)$ is a $q \times n$ matrix of subject-level covariates; $B = (\beta_1, \ldots, \beta_n)$, and "$\otimes$" is the Kronecker product.

3. Generate Λ^{-1} from the Wishart distribution $W_p(\eta_n, \Delta_n)$ where $\eta_n = \eta_0 + n$, and $\Delta_n = (\Delta_0^{-1} + \Sigma_{i=1}^n (\beta_i - \Theta z_i)(\beta_i - \Theta z_i)')^{-1}$.

4. Independently generate σ_1^2 to σ_n^2 from inverse gamma distributions: $\sigma_i^2 \sim IG(\alpha_i/2, \psi_i/2)$ where $\alpha_i = \alpha + m_i$, and $\psi_i = \psi + (Y_i - X_i\beta_i)'(Y_i - X_i\beta_i)$.

5. Generate ψ from the gamma distribution $G(r_n/2, s_n/2)$ where $r_n = r_0 + n\alpha$, and $s_n = s_0 + \Sigma_{i=1}^n \sigma_i^{-2}$.

6. The generation of α given the data and the other parameters is the only non-standard component of the analysis of the model. First, we approximate the density of α given the current values of $\{\sigma_i^2\}$ by a lognormal distribution where the mode $\tilde{\alpha}$ and curvature C match. Then, the approximating lognormal distribution is combined with the prior distribution for α, which is also a lognormal distribution, to obtain the approximating posterior distribution. A candidate value for α is generated from this approximating posterior distribution: $\log(\alpha)$ is $N(a_n, d_n^2)$ where $d_n^2 = (d_0^{-2} + \tau^{-2})^{-1}$, $a_n = d_n^2(\tau^{-2}\mu + d_0^{-2}a_0)$, $\tau^2 = -(\tilde{\alpha}^2C)^{-1}$, and $\mu = \log(\tilde{\alpha}) + \tau^2$. The Metropolis algorithm (Hastings 1970, Smith and Roberts 1993, and Tanner 1993) is used to accept or reject the candidate value. The Metropolis algorithm is a general method of compensating for the fact that the candidate for α was generated from an approximating distribution. Simulation studies of this procedure indicate that it generates random deviates from the posterior distribution of α.

The Markov chain Monte Carlo (MCMC) initialized all partworths, Θ, and covariances to zero and all variances to one. The MCMC ran for 11,000 iterates, of which the last 10,000 were used to estimate posterior means and variances.

Appendix B

Fisher's Information

The first step in computing the information matrix is to obtain the log-likelihood of the data given the parameters. Because we are interested in Θ and Λ, we integrate $\{\beta_i\}$ out of the model to obtain: $Y_i = X_i\Theta z_i + \epsilon_i^*$ where ϵ_i^* has a multivariate normal distribution with mean 0 and covariance $\Sigma_i = \sigma_i^2 I_{m_i} + X_i\Lambda X_i'$. Because the error terms are mutually independent, so are the $\{\epsilon_i^*\}$. The mean of Y_i can be expressed as: $X_i\Theta z_i = (z_i' \otimes X_i)\,\text{vec}(\Theta)$ where "$\otimes$" is the Kronecker product.

The log-likelihood of Θ and Λ is:

$$\log[f(Y|\Theta, \Lambda)] = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\log|\Sigma_i|$$

$$-\frac{1}{2}\sum_{i=1}^{n}[Y_i - (z_i' \otimes X_i)\Theta^v]'\Sigma_i^{-1}[Y_i - (z_i' \otimes X_i)\Theta^v],$$

where $T = \Sigma_{i=1}^n m_i$. Because we will compare designs with the same prior distributions, we will not include the information of the prior distributions.

The following lemma collects well-known facts from linear algebra regarding derivatives.

LEMMA 1. *Let x be a J-dimensional vector, and let A and B be $I \times J$ and $J \times J$ matrices that do not depend on x. Then $\frac{\partial}{\partial x}Ax = A'$, and $\frac{\partial}{\partial x}x'Bx = 2Bx$. If A is a nonsingular, $J \times J$ matrix, then*

$$\frac{\partial A^{-1}}{\partial x} = -A^{-1}\frac{\partial A}{\partial x}A^{-1} \quad and \quad \frac{\partial \log|A|}{\partial A} = 2A^{-1} - \mathrm{diag}(A^{-1})$$

if A is symmetric and A^{-1} otherwise.

PROOF. See Timm (1975, p. 97–103).

The next lemma computes derivatives needed for the Fisher's information.

LEMMA 2. *If Λ is a symmetric $p \times p$ matrix and if X is a $J \times p$ matrix with column vectors: $X = [x_1, \ldots, x_p]$, then*

$$\frac{\partial}{\partial \lambda_{u,v}} X\Lambda X' = \begin{cases} x_u x_u' & \text{if } u = v, \\ x_u x_v' + x_v x_u' & \text{if } u \neq v. \end{cases} \tag{9}$$

Define $\Sigma = \sigma^2 I + X\Lambda X'$. Then

$$\frac{\partial \log(|\Sigma|)}{\partial \lambda_{u,v}} = \mathrm{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \lambda_{u,v}}\right) = \begin{cases} x_u'\Sigma^{-1}x_u & \text{if } u = v, \\ 2x_u'\Sigma^{-1}x_v & \text{if } u \neq v. \end{cases} \tag{10}$$

PROOF. Use $X\Lambda X' = \Sigma_{i,j}\lambda_{i,j}x_i x_j'$ to obtain Equation (9). To show Equation (10), let $\sigma_{i,j}$ and $\sigma^{i,j}$ be the i, j elements of Σ and Σ^{-1}, respectively. Both Σ and Σ^{-1} are symmetric. Using the chain rule,

$$\frac{\partial \log|\Sigma|}{\partial \lambda_{u,v}} = \sum_{i=1}^{J}\sum_{j=i}^{J}\frac{\partial \log|\Sigma|}{\partial \sigma_{i,j}}\frac{\partial \sigma_{i,j}}{\partial \lambda_{u,v}}$$

$$= \sum_{i=1}^{J}\sigma^{i,i}\frac{\partial \sigma_{i,i}}{\partial \lambda_{u,v}} + \sum_{i=1}^{J-1}\sum_{j=i+1}^{J}2\sigma^{i,j}\frac{\partial \sigma_{i,j}}{\partial \lambda_{u,v}} = \mathrm{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \lambda_{u,v}}\right).$$

PROOF OF THEOREM 1. Define $r_i = Y_i - (z_i' \otimes X_i)\Theta^v$. The partial derivatives of the marginal density of Y with respect to Θ are

$$\frac{\partial \log f(Y|\Theta, \Lambda)}{\partial \Theta^v} = \sum_{i=1}^{n}(z_i \otimes X_i')\Sigma_i^{-1}r_i,$$

$$\frac{\partial^2 \log f(Y|\Theta, \Lambda)}{\partial(\Theta^v)^2} = -\sum_{i=1}^{n}(z_i \otimes X_i')\Sigma_i^{-1}(z_i' \otimes X_i).$$

The Fisher's information for Θ is

$$FI(\Theta, \Theta) = \sum_{i=1}^{n}(z_i \otimes X_i')\Sigma_i^{-1}(z_i' \otimes X_i) = \sum_{i=1}^{n}(z_i z_i' \otimes X_i'\Sigma_i^{-1}X_i),$$

where the last line results from z_i being a vector. $FI(\Theta, \Lambda)$ is computed from

$$\frac{\partial^2 \log f(Y|\Theta, \Lambda)}{\partial \Theta^v \partial \lambda_{u,v}} = -\sum_{i=1}^{n}(z_i \otimes X_i')\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \lambda_{u,v}}\Sigma_i^{-1}r_i.$$

Using $E(r_i) = 0$, $FI(\Theta, \Lambda) = FI(\Lambda, \Theta)' = 0$. The information matrix for Λ results from the following computations:

$$\frac{\partial f(Y|\Theta, \Lambda)}{\partial \lambda_{u,v}} = -\frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\left(\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \lambda_{u,v}}\right) + \frac{1}{2}\sum_{i=1}^{n}r_i'\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \lambda_{u,v}}\Sigma_i^{-1}r_i,$$

$$\frac{\partial^2 f(Y|\Theta, \Lambda)}{\partial \lambda_{u,v}\partial \lambda_{r,s}} = \frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\left(\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \lambda_{r,s}}\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \lambda_{u,v}}\right)$$

$$-\frac{1}{2}\sum_{i=1}^{n}r_i'\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \lambda_{r,s}}\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \lambda_{u,v}}\Sigma_i^{-1}r_i$$

$$-\frac{1}{2}\sum_{i=1}^{n}r_i'\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \lambda_{u,v}}\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \lambda_{r,s}}\Sigma_i^{-1}r_i,$$

$$FI(\lambda_{u,v}, \lambda_{r,s}) = \frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\left(\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \lambda_{u,v}}\Sigma_i^{-1}\frac{\partial \Sigma_i}{\partial \lambda_{r,s}}\right).$$

Using Equation (9) for the different values of u, v, r, and s results in the theorem.

PROOF OF COROLLARY 1. $X'\Sigma^{-1}X = m/(\sigma_0^2 + \lambda_0^2 m)I$. The Fisher's information for Θ from Theorem 1 is

$$FI(\Theta, \Theta) = Z'Z \otimes \frac{m}{\sigma_0^2 + \lambda_0^2 m}I.$$

The Fisher's information for the variance Λ is

$$FI(\lambda_{u,u}, \lambda_{u,u}) = \frac{1}{2}n\left(\frac{m}{\sigma_0^2 + \lambda_0^2 m}\right)^2,$$

$$FI(\lambda_{u,v}, \lambda_{u,v}) = n\left(\frac{m}{\sigma_0^2 + \lambda_0^2 m}\right)^2 \quad \text{if } u \neq v,$$

$$FI(\lambda_{u,v}, \lambda_{r,s}) = 0 \text{ otherwise.}$$

The logarithm of the determinant of the Fisher's information is computed by noting that if A is a $q \times q$ matrix and if B is a $p \times p$ matrix, then $|A \otimes B| = |A|^p|B|^q$.

PROOF OF THEOREM 2. Define $g(n, m) = \log|FI| - p\log|n^{-1}Z'Z|$ where $\log|FI|$ is given by Equation (4). Assuming that n and m are real numbers, and not just positive integers, the cost constraint becomes an equality. Solve for the number of subjects: $n = c_T/(c + dm)$; substitute into $g(n, m)$; and differentiate to obtain

$$\frac{p(p + q + 1)\sigma_0^2}{m(\sigma_0^2 + \lambda_0^2 m)} - \frac{p(p + 2q + 1)d}{2(c + dm)}.$$

The maximizing value of m, without taking into account that the m is a positive integer such that the design is orthogonal, is m^* given by Equation (6). Because the criterion function is continuous with a global maximum at m^* for positive values of m, the maximizing value of m, when restricted to feasible values, will either occur at one of the endpoints of the feasible region or at one of the two integer values of the feasible region that are closest to m^*. The optimal m will be at the lower endpoint, L, if $m^* \leq L$ and at the upper endpoint, U, if $m^* \geq U$. Rewriting these inequalities in terms of λ_0^2/σ_0^2 gives the theorem.

The proof of Theorem 3 uses the next lemma.

LEMMA 3. *Suppose that prior estimates of the variances are σ_0^2 and $\Lambda = \lambda_0^2 I$. Let X_i be one of the blocks from a 2^{f-b} blocked factorial design. Then*

$$X_i'\Sigma_i^{-1}X_i = \frac{2^{-b}}{\sigma_0^2/2^f + \lambda_0^2}\,\text{block}(P_i, \ldots, P_i).$$

PROOF. From the inverse of partitioned matrices,

$$X_i'\Sigma_i^{-1}X_i = \sigma_0^{-2}X_i'X_i - \sigma_0^{-2}X_i'X_i(\lambda_0^{-2}I + \sigma_0^{-2}X_i'X_i)^{-1}\sigma_0^{-2}X_i'X_i.$$

Because $X_i'X_i = 2^{f-b}$ block $(P_i, \ldots, P_i)$, we have $\lambda_0^{-2}I + \sigma_0^{-2}X_i'X_i$ is a block diagonal matrix with $\lambda_0^{-2}I + \sigma_0^{-2}2^{f-b}P_i$ along the diagonal. Using $P_i^2 = 2^bP_i$ (John 1969), it is easy to verify that:

$$(aI + cP_i)^{-1} = a^{-1}I - \frac{c}{a(a + c2^b)}P_i$$

by multiplication. Hence,

$$(\lambda_0^{-2}I + \sigma_0^{-2}2^{f-b}P_i)^{-1} = \lambda_0^2 I - \frac{2^{f-b}\lambda_0^4}{\sigma_0^2 + 2^f\lambda_0^2}P_i.$$

Using $P_i^2 = 2^bP_i$, shows that $X_i'\Sigma_i^{-1}X_i$ is a block diagonal matrix with blocks:

$$2^{f-b}\sigma_0^{-2}P_i - 2^{2f-2b}\sigma_0^{-4}P_i\left(\lambda_0^2 I - \frac{2^{f-b}\lambda_0^4}{\sigma_0^2 + 2^f\lambda_0^2}P_i\right)P_i = \frac{2^{f-b}}{\sigma_0^2 + 2^f\lambda_0^2}P_i,$$

PROOF OF THEOREM 3. Let z_k be the common covariate for all of the subjects in group k. For subjects in block j and group k, define $\Sigma_{j,k} = \sigma_{0,k}^2 I + \lambda_0^2 X_{j,k}X_{j,k}'$. Then

$$FI(\Theta, \Theta) = \sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{2^{b_k}} z_k z_k' \otimes X_{j,k}'\Sigma_{j,k}^{-1}X_{j,k}$$

$$= \sum_{k=1}^{q} z_k z_k' n_k 2^{-b_k}a_k \otimes \sum_{j=1}^{2^{b_k}} \text{block}(P_j, \ldots, P_j).$$

Because the full design is orthogonal, i.e., $X'X = 2^fI$, and because $X'X = \sum_{i=1}^{2^b} X_i'X_i$, we see that $\sum_{j=1}^{2^b} P_j = 2^bI$. Thus, we have

$$FI(\Theta, \Theta) = \left(\sum_{k=1}^{q} n_k a_k z_k z_k'\right) \otimes I.$$

A similar argument gives $FI(\lambda_{u,u}, \lambda_{u,u})$.

PROOF OF THEOREM 4. The Lagrangian is

$$g(n, b) = v\sum_{k=1}^{q}\log(n_k a_k) + w\log\left(\frac{1}{2}\sum_{k=1}^{q} n_k 2^{-b_k}a_k^2\right)$$

$$- \gamma\left(\sum_{k=1}^{q} n_k 2^{b_k}[c_k + d_k(2^{f-b_k})] - c_T\right).$$

Given $b_1, \ldots, b_q$, the optimal n_j satisfies

$$\frac{\partial g(n, b)}{\partial n_j} = \frac{v}{n_j} + \frac{w}{2}2^{-b_j}a_j^2\left(\frac{1}{2}\sum_{k=1}^{q} n_k 2^{-b_k}a_k^2\right)^{-1} - \gamma 2^{b_j}[c_j + d_j(2^{f-b_j})] = 0.$$

Multiplying the above partial derivative by n_j, summing over j, and solving for γ results in $\gamma = (vq + w)/c_T$.

References

Adams, L. L. and D. Gale (1982), "Solving the Quandary between Questionnaire Length and Response Rate in Educational Research," *Research in Higher Education*, 17(3), 231–240.

Allenby, G. M., N. Arora, and J. L. Ginter (1995), "Incorporating Prior Knowledge into the Analysis of Conjoint Studies," *Journal of Marketing Research*, 32, 152–162.

—— and J. L. Ginter (1994), "Hierarchical Conjoint Analysis," Working Paper Series, Max Fisher College of Business, The Ohio State University, WPS 94-5.

—— and P. J. Lenk (1994), "Modeling Household Purchase Behavior with Logistic Normal Regression," *Journal of the American Statistical Association*, 89, 1218–1229.

—— and —— (1995), "Reassessing Brand Loyalty, Price Sensitivity, and Merchandising Effects on Consumer Brand Choice," *Journal of Business & Economic Statistics*, 13, 281–298.

Arnold, S. F. (1993), "Gibbs Sampling," in *Handbook of Statistics*, 9, C. R. Rao (Ed.), Amsterdam: North-Holland, 599–626.

Bean, A. G. and M. J. Roszkowski (1995), "The Long and Short of It: When Does Questionnaire Length Affect Response Rate," *Marketing Research*, 7, 21–26.

Berdie, D. R. (1989), "Reassessing the Value of High Response Rates to Mail Surveys," *Marketing Research: A Magazine of Management and Applications*, 1.

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag.

Biner, P. M. (1988), "Effects of Cover Letter Appeal and Monetary Incentives on Survey Response: A Reactance Theory Application," *Basic & Applied Social Psychology*, 9(2), 99–106.

—— and D. L. Barton (1990), "Justifying the Enclosure of Monetary Incentives in Mail Survey Cover Letters," *Psychology and Marketing*, 7(3), 153–162.

—— and H. J. Kidd (1994), "The Interactive Effects of Monetary Incentive Justification and Questionnaire Length on Mail Survey Response Rates," *Psychology and Marketing*, 11(5), 483–492.

Brennan, M. (1992), "The Effect of a Monetary Incentive on Mail Survey Response Rates: New Data," *Journal of the Market Research Society*, 34(2), 173–177.

Bretton-Clark (1988), *Conjoint Analysis Software*, Morristown, NJ: Bretton-Clark, Inc.

Brown, T. L., D. J. Decker, and N. A. Connelly (1989), "Response to Mail Surveys on Resource-based Recreation Topics: A Behavioral Model and Empirical Analysis," *Leisure Sciences*, 11, 99–110.

Box, G. E. P. and G. C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Reading, Massachusetts: Addison-Wesley Publishing.

Cattin, P. and D. Wittink (1982), "Commercial Use of Conjoint Analysis: A Survey," *Journal of Marketing*, 46, 44–53.

Childers, T. L. and O. C. Ferrel (1979), "Response Rates and Perceived Questionnaire Length in Mail Surveys," *Journal of Marketing Research*, 16(3), 429–431.

DeGroot, M. H. (1970), *Optimal Statistical Decisions*, New York: McGraw-Hill.

DeSarbo, W. S., D. J. Howard, and K. Jedidi (1990), "MULTICLUS: A New Method for Simultaneously Performing Multidimensional Scaling and Cluster Analysis," *Psychometrika*, 121–136.

DeSarbo, W. S., R. L. Oliver, and A. Rangaswamy (1989), "A Simulated Annealing Methodology of Clusterwise Linear Regression," *Psychometrika*, 54(4), 707–736.

DeSarbo, W. S., M. Wedel, M. Vriens, and V. Ramaswamy (1992), "Latent Class Metric Conjoint Analysis," *Marketing Letters*, 3, 273–288.

Dillman, D. A. (1991), "The Design and Administration of Mail Surveys," *Annual Review of Sociology*, 17, 225–249.

——, M. D. Sinclair, and J. R. Clark (1993), "Effects of Questionnaire Length, Respondent-friendly Design, and a Difficult Question on Response Rates for Occupant-addressed Census Main Surveys," *Public Opinion Quarterly*, 57, 289–304.

Dong, H.-K. (1983), "Method of Complete Triads: An Investigation of Unreliability in Multidimensional Perception of Nations," *Multivariate Behavior Research*, 18, 85–96.

Gelfand, A. E. and A. F. M. Smith (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Green, P. E. and A. M. Krieger (1994a), "Conjoint Analysis with Product—Positioning Applications," in J. Eliashberg and G. Lilien (Eds.), *Marketing*, 5, New York: North-Holland.

—— and —— (1994b), "Hybrid Models in Conjoint Analysis: An Update," Working Paper, Wharton School, University of Pennsylvania, January.

——, ——, and M. K. Agarwal (1991), "Adaptive Conjoint Analysis: Some Caveats and Suggestions," *Journal of Marketing Research*, 28, 215–222.

——, ——, and C. M. Schaffer (1993), "A Hybrid Conjoint Model with Individual Level Interaction," *Advances in Consumer Research*, 20, 1–6.

—— and V. R. Rao (1971), "Conjoint Measurement for Quantifying Judgmental Data," *Journal of Marketing Research*, 8, 355–363.

—— and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research*, 5, 103–123.

—— and —— (1990), "Conjoint Analysis in Marketing Research: New Developments and Directions," *Journal of Marketing*, 54, 3–19.

Gunn, W. J. and I. N. Rhodes (1981), "Physician Response Rates to a Telephone Survey: Effects of Monetary Incentive Level," *Public Opinion Quarterly*, 45(1), 109–115.

Hagerty, M. R. (1985), "Improving the Predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis," *Journal of Marketing Research*, 22, 168–184.

Hansen, R. A. (1980), "A Self-perception Interpretation of the Effect of Monetary and Nonmonetary Incentives on Mail Survey Respondent Behavior," *Journal of Marketing Research*, 17(1), 77–83.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Application," *Biometrika*, 57, 97–109.

Heberlein, T. A. and R. Baumgartner (1978), "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature," *American Sociological Review*, 43(4), 447–462.

Helgeson, J. G. and M. L. Ursic (1993), "Information Load, Cost/Benefit Assessment and Decision Strategy Variability," *Journal of the Academy of Marketing Science*, 21(1), 13–20.

Herman, S. (1988), "Software for Full-profile Conjoint Analysis," in *Proceeding of the Sawtooth Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, Ketchum ID, Sawtooth Software.

Herzog, A. R. and J. G. Bachman (1981), "Effects of Questionnaire Length on Response Quality," *Public Opinion Quarterly*, 45(4), 549–559.

Hill, B. M. (1965), "Inference about Variance Components in the One-way Model," *Journal of the American Statistical Society*, 60, 806–825.

Huck, S. W. and E. M. Gleason (1974), "Using Monetary Inducements to Increase Response Rates from Mailed Surveys: A Replication and Extension of Previous Research," *Journal of Applied Psychology*, 59(2), 222–225.

Jacoby, J. (1977), "Information Load and Decision Quality: Some Contested Issues," *Journal of Marketing Research*, 14, 569–573.

—— (1984), "Perspectives on Information Overload," *Journal of Consumer Research*, 10(4), 432–435.

——, D. E. Speller, and C. K. Berning (1974), "Brand Choice Behavior as a Function of Information Load: Replication and Extension," *Journal of Consumer Research*, 1(1), 33–42.

——, ——, and C. A. Kohn (1974), "Brand Choice Behavior as a Function of Information Load," *Journal of Marketing Research*, 11, 63–69.

——, G. J. Szybillo, and J. Bustao-Schach (1977), "Information Acquisition Behavior in Brand Choice Situations," *Journal of Consumer Research*, 3(4), 209–216.

Jedidi, K., R. Kohli, and W. S. DeSarbo (1995), "Consideration/choice Sets in Conjoint Analysis: A Latent Structure Approach," Working Paper, Columbia University, New York, NY.

John, P. W. M. (1969), "Some Non-orthogonal Fractions of 2^n Designs," *Journal of the Royal Statistical Society, Series B*, 31, 270–275.

—— (1971), *Statistical Design and Analysis of Experiments*, New York: Macmillan.

Johnson, R. D. (1987), "Adaptive Conjoint Analysis Software," in *Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, Ketchum, ID; Sawtooth Software, 253–265.

—— (1993), "Choice-based Conjoint Analysis Software," Sawtooth Software.

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

Kamakura, W. A. (1988), "A Least Squares Procedure for Benefit Segmentation with Conjoint Experiments," *Journal of Marketing Research*, 25, 157–167.

Keller, K. L. and R. Staelin (1987), "Effects of Quality and Quantity on Decision Effectiveness," *Journal of Consumer Research*, 14, 200–213.

—— and —— (1989), "Assessing Biases in Measuring Decision Effectiveness and Information Overload," *Journal of Consumer Research*, 15, 504–508.

Kraut, A. I., A. D. Wolfson, and A. Rothenberg (1975), "Some Effects of Position on Opinion Survey Items," *Journal of Applied Psychology*, 60(6), 774–776.

Lee, J.-W. and J. F. Yates (1992), "How Quantity Judgement Changes as the Number of Cues Increases: An Analytical Framework and Review," *Psychological Bulletin*, 112(2), 363–377.

Lenk, P. (1992), "Hierarchical Bayes Forecasts of Multinomial Dirichlet Data Applied to Coupon Redemptions," *Journal of Forecasting*, 11, 603–619.

—— and A. G. Rao (1990), "New Models from Old: Forecasting Product Adoption by Hierarchical Bayes Procedures" with discussion, *Marketing Science*, 9, 42–57.

Lindley, D. V. and A. F. M. Smith (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society, Series B*, 34, 1–41.

Lockhart, D. C. (1990), "The Effect of Incentives on the Return Rate of Surveys to Physicians," *Proceedings of the Society for Consumer Psychology*, 122–124.

Louviere, J. J. (1988), *Analyzing Decision Making: Metric Conjoint Analysis*, Beverly Hills, CA: Sage Publications.

Malhotra, N. K. (1982), "Information Load and Consumer Decision Making," *Journal of Consumer Research*, 8(4), 419–430.

—— (1984), "Reflections on the Information Overload Paradigm in Consumer Decision Making," *Journal of Consumer Research*, 10(4), 436–440.

—— (1986), "An Approach to the Measurement of Consumer Preferences Using Limited Information," *Journal of Marketing Research*, 23, 33–49.

——, A. K. Jain, and S. W. Lagakos (1982), "The Information Overload Controversy: An Alternative Viewpoint," *Journal of Marketing*, 46(2), 27–37.

Meyer, R. and E. Johnson (1989), "Information Overload and the Nonrobustness of Linear Models: A Comment on Keller and Staelin," *Journal of Consumer Research*, 15, 498–503.

Paolillo, J. G. and P. Lorenzi (1984), "Monetary Incentives and Mail Questionnaire Response Rates," *Journal of Advertising*, 13(1), 46–48.

Payne, J. W. (1976), "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," *Organizational Behavior and Human Performance*, 16, 366–387.

Pilz, J. (1991), *Bayesian Estimation and Experimental Design in Linear Regression Models*, New York: John Wiley & Sons.

Rao, V. R. (1977), "Conjoint Measurement in Marketing Analysis," in *Multivariate Methods for Market and Survey Research*, J. N. Sheth ed. American Marketing Association, Chicago, 257–286.

Rossi, P. E. and G. M. Allenby (1993), "A Bayesian Approach to Estimating Household Parameters," *Journal of Marketing Research*, 30, 171–182.

Roszkowski, M. J. and A. G. Bean (1990), "Believe It or Not! Longer Questionnaires Have Lower Response Rates," *Journal of Business and Psychology*, 4(4), 495–509.

Smith, A. F. M. (1973), "A General Bayesian Linear Model," *Journal of the Royal Statistical Society, Series B*, 35, 67–75.

—— and G. O. Roberts (1993), "Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society Series B*, 55, 3–23.

Srinivasan, V., A. K. Jain, and N. K. Malhotra (1983), "Improving the Predictive Power of Conjoint Analysis by Constrained Parameter Estimation," *Journal of Marketing Research*, 20, 433–438.

Summers, J. O. (1974), "Less Information Is Better?" *Journal of Marketing Research*, 11, 467–468.

Tanner, M. A. (1993), *Tools for Statistical Inference*, Lecture Notes in Statistics 67, New York: Springer-Verlag.

—— and W. H. Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Society*, 82, 528–540.

Timm, N. H. (1975), *Multivariate Analysis with Applications in Education and Psychology*, Monterey, California: Brooks/Cole Publishing.

Van der Lans, I. A. and W. H. Heiser (1992), "Constrained Part-worth Estimation in Conjoint Analysis Using the Self-explicated Utility Model," *International Journal of Research in Marketing*, 9, 325–344.

Wedel, M. and J.-B. E. M. Steenkamp (1989), "Fuzzy Clusterwise Regression Approach to Benefit Segmentation," *International Journal of Research in Marketing*, 6, 241–258.

Wilkie, W. L. (1974), "Analysis of Effects of Information Load," *Journal of Marketing Research*, 11, 462–466.

Wittink, D. R. and P. Cattin (1989), "Commercial Use of Conjoint Analysis: An Update," *Journal of Marketing*, 53, 91–96.

——, M. Vriens, and W. Burhenne (1994), "Commercial Use of Conjoint Analysis in Europe: Results and Critical Reflections," *International Journal of Research in Marketing*, 11, 1, 41–52.

Wright, P. (1974), "The Harassed Decision Maker: Time Pressures, Distractions, and the Use of Evidence," *Journal of Applied Psychology*, 59(5), 555–561.

Yu, J. and H. Cooper (1983), "A Quantitative Review of Research Design Effects on Response Rates to Questionnaires," *Journal of Marketing Research*, 20, 36–44.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York: Wiley & Sons.

This paper was received April 5, 1994, and has been with the authors 326 days for 3 revisions; processed by Russell Winer.