Efficient Experimental Designs for Hyper-parameter Estimation: Studying the Level-Effect in Conjoint Analysis

by

Qing Liu
Department of Statistics
Ohio State University
qliu@stat.ohio-state.edu

Angela Dean
Department of Statistics
Ohio State University
amd@stat.ohio-state.edu

David Bakken
Harris Interactive
DBakken@HarrisInteractive.com

Greg M. Allenby
Fisher College of Business
Ohio State University
allenby.1@osu.edu

December, 2007

Efficient Experimental Designs for Hyper-parameter Estimation: Studying the Level-Effect in Conjoint Analysis

Abstract

Research in marketing, and business in general, involves understanding when effect-sizes are expected to be large and when they are expected to be small. An example is the level-effect in marketing, where the effect of product attributes on utility is positively related to the number of levels present among choice alternatives. Knowing when consumers are sensitive to the levels of attributes is an important aspect of merchandising, selling and promotion. In this paper, we propose efficient methods of learning about contextual factors that influence consumer preference and sensitivities within the context of a hierarchical Bayes model. A design criterion is developed for hierarchical linear models, and validated in a study of the "level-effect" in conjoint analysis using a national sample of respondents. Extensions to other model structures are discussed.

Efficient Experimental Designs for Hyper-parameter Estimation: Studying the Level-Effect in Conjoint Analysis

1. Introduction

The level-effect in conjoint analysis was first observed by Currim, Weinberg and Wittink (1981), who reported an increase in the estimated importance of attribute-levels as the number of intermediate levels increased. Since then, the level-effect has been observed across different measurement scales, various data collection methods and estimation techniques (see Wittink et al., 1982, 1990, 1997; Steenkamp and Wittink, 1994; and Creyer and Ross, 1988). The existence of the level-effect suggests that consumer preference varies with context. Specifically, it implies that, in the market place, consumer preference sensitivity to a product attribute is affected by the variety of attribute levels displayed, which varies from store to store. Such a finding is consistent with a large body of consumer behavior literature (e.g., Huber, Payne and Puto, 1982; Lynch, Chakravarti and Mitra, 1991; Simonson and Tversky, 1992) describing the influence of contextual effects in choice. Standard conjoint analysis models in marketing research do not adjust for contextual influences in part-worth estimation and, consequently, cannot predict consumer buying behavior well in different contexts.

The study of contextual effects is common in business research. Examples range from studies within a firm (e.g., the relative effectiveness of alternative employee incentives) to studies of market-level behavior, such as the impact of macro-economic variables on financial markets, the role of consumer confidence on retail spending, and variation in the effectiveness of marketing's four P's (i.e., product, price, place and promotion). The study of effect-size variation has important implications for resource allocation, including the timing of offers, competitive effects, and any decision that can be implemented in a customized, non-uniform manner.

The accurate estimation of the level-effect, and contextual effects in general, requires models that describe the behavior of individual-level units of analysis, such as the customers of a firm, and their across-unit variation. The units can be described in terms of unit-specific variables, such as demographics, those that describe the context of the choice environment, or both. Hierarchical models are often used in such analysis, with the hyper-parameters in the upper-level of the model describing intra-unit variation in effects. The accurate estimation of hyper-parameters is critical for knowing when effect-sizes are expected to be large.

This paper investigates experimental designs for accurate hyper-parameter estimation in hierarchical linear models. The accurate estimation of hyper-parameters involves the consideration of two aspects of design choice: i) the within-unit design specification and ii) the across-unit design specification. These specifications interact in hierarchical models, and can lead to optimal designs that are non-orthogonal. Thus, designs for the accurate estimation of respondent-level-effects alone may not be efficient for hyper-parameter estimation, as demonstrated in Section 4 through the comparison of the survey designs used in our study of the level-effect.

We develop and explore a simultaneous design criterion for within- and across-unit specifications. We note that our criterion also applies to simpler versions of the general model. When faced with a fixed population of units (e.g., respondents), the across-unit design is fixed and the analyst is faced with the task of selecting the best within-unit design (e.g., product concepts in a conjoint analysis) for hyper-parameter estimation. Alternatively, there are situations where the within-unit design is fixed and the goal is to determine the best across-unit design. An example is the testing a pre-determined set of product offerings, with the goal of identifying the best set of respondents to survey.

We illustrate the performance of our design criterion in a study of the level-effect in conjoint analysis. As noted earlier, the level-effect is a positive association between estimated effect sizes and the number of attribute levels present in a conjoint study. We propose a model for the level-effect that contains a functional relationship between the within-unit specification, where respondents evaluate products described in terms of attributes and their levels, and the across-unit specification described in terms of the relative locations, ranks and the number of attribute-levels studied. We find evidence for the level-effect and demonstrate the usefulness of the proposed design criterion.

This paper contributes to the literature in two ways. First, we call attention to the importance of hyper-parameter estimation. While hierarchical models have incorporated hyper-parameters as coefficients of regressors in many marketing studies, not much attention has been given to methods for their accurate estimation, nor their importance in guiding business practice. We present a general design criterion for estimating these parameters and illustrate its use in a practical setting. Second, we develop and test a structural model for the level-effect in conjoint analysis. This effect can only be studied with hyper-parameters in a hierarchical model because the same respondent cannot be exposed practically to multiple choice contexts in a conjoint study. Inferences about the level-effect must therefore come from an across-unit analysis. Our model parameterizes the level-effect so that counterfactual predictions can be made to new contexts, as opposed to employing a dummy variable coding of the contexts that does not allow for such predictions. We show that our model of the level-effect fits the data better than simpler specifications.

The remainder of the paper is organized as follows: section 2 describes the optimal design criterion for the estimation of hyper-parameters in a hierarchical linear model. Section 3 introduces the theory and parameterization for studying the level-effect in conjoint analysis. Designs are constructed and compared in section 4, and section 5 reports empirical results from a national web-based survey. Concluding remarks are offered in section 6.

2. Design Criterion for Hyper-parameters

Consider a consumer survey in which respondent i (i=1, …, n) is presented with a set of m_i profiles (stimuli). The stimuli are constructed with information on various levels of marketing variables, such as price, product attributes or possibly aspects of advertisements. The levels of the marketing variables are reflected in the $m_i \times p$ model matrix, X_i. The responses of consumer i to the set of stimuli are represented by the vector y_i of length m_i. The effects of the various levels of marketing variables on respondent i's responses are captured by the p elements of vector β_i, which are assumed to be random effects distributed according to a multivariate normal distribution with mean $Z_i\theta$ and variance-covariance matrix Λ (p × p). Z_i is a matrix (p × q) of covariates, such as household income, age or other contextual variables that characterize the purchasing environment. θ is a parameter vector of length q. Thus, the hierarchical linear model is of the following form:

$$y_i \mid \beta_i, \sigma^2 = \mathrm{X}_i \beta_i + \varepsilon_i \qquad (1)$$

$$\beta_i \mid \theta, \Lambda = \mathrm{Z}_i \theta + \delta_i \qquad (2)$$

The error vector, ε_i of length m_i in the first level of the hierarchy captures consumer i's response variability to the set of stimuli, and it is assumed to have a Multivariate Normal distribution with mean vector 0 of length m_i and variance-covariance matrix $\sigma^2 \mathrm{I}_{m_i}$. The error vector δ_i of length p in the second level of the hierarchy captures the dispersion of the individual-level-effects β_i and is assumed to be Multivariate Normal with mean vector 0 and variance-covariance matrix Λ of size p x p. When the prior knowledge is weak, the following proper but diffuse priors are usually assumed for θ, Λ and σ^2 (see Gamerman 1997; Rossi, Allenby and McCulloch 2005). These are replaced by more informative priors when prior knowledge is available.

$$\theta \sim \text{Normal}\,(0, 100I_p) \tag{3}$$

$$\Lambda \sim \text{Inverted Wishart}\,(\nu_0 = p+3, V_0 = \nu_0 I_p) \tag{4}$$

$$\sigma^2 \sim \text{Inverse Gamma}\,(3/2, 1/2) \tag{5}$$

Accurate estimation of hyper-parameters is important in situations when learning of contextual effects is of interest, or when predictions to new contexts are required. Pragmatic approaches to finding efficient designs have been proposed for the estimation of hyper-parameters under a hierarchical nonlinear model. For example, the swapping, relabeling and cycling heuristic by Sándor and Wedel (2002); the linearization approach by Mentré et al. (1997); the stochastic gradient search by Tod et al. (1998), and the ``MCMC nested within Monte Carlo" approach by Han and Chaloner (2004). No contextual variables are included in the hierarchical models used in these papers, that is, Z_i equals the identity matrix.

Under a hierarchical linear model with independent, homoscedastically distributed random effects (i.e., $\Lambda = \lambda^2 I$), Lenk et al. (1996) analytically investigate, in the survey setting, the tradeoff between the number of subjects and the number of questions per subject under a cost constraint and an orthogonal design structure. Contextual variables are included in the model but Z_i is assumed given, that is, Z_i is not under the control of the experimenter. In this paper, we explore efficient experimental designs expressed in terms of the stimuli (X_i) and contextual variables (Z_i).

Design Optimality Criterion for Hyper-parameter Estimation

We consider the situation when the primary interest is on the accurate estimation of the context effect vector θ, or a function $g(\theta)$ of vector θ, while the response error variance σ^2 and the random effects variance-covariance matrix Λ are treated as nuisance parameters. Equations (1) and (2) can be combined to obtain

$$y_i \mid \theta, \Lambda, \sigma^2 \sim \mathrm{N}_{m_i}\left(\mathrm{X}_i \mathrm{Z}_i \theta, \quad \Sigma_i = \sigma^2 \mathrm{I}_{m_i} + \mathrm{X}_i \Lambda \mathrm{X}_i^{'}\right), \tag{6}$$

where N_{m_i} denotes multivariate normal distribution with dimension m_i, with priors (3) to (5) assumed for θ, Λ and σ^2.

We define an optimal design for the context effect vector θ is a design with matrices $\{X_i\}$ and $\{Z_i\}$ (if Z_i can be controlled by the experimenter), i=1, …, n, that maximizes the expected gain in Shannon information (see Chaloner and Verdinelli, 1995, page 277):

$$\psi = \int \left\{ \log \left| \sum_{i=1}^{n} \mathrm{Z}_i^{'} \mathrm{X}_i^{'} (\sigma^2 \mathrm{I}_{m_i} + \mathrm{X}_i \Lambda \mathrm{X}_i^{'})^{-1} \mathrm{X}_i \mathrm{Z}_i \right| \right\} p(\Lambda) p(\sigma^2) d\Lambda d\sigma^2, \tag{7}$$

where p(Λ) and p(σ^2) denote the probability density functions of the prior distributions of Λ and σ^2. Details of the derivation of the ψ-criterion in (7) can be found in Liu et. al. (2007). We note that the ψ-criterion is essentially the Bayesian D-criterion for the estimation of hyper-parameter vector θ. For notational convenience, we call it the ψ-criterion to distinguish it from the Bayesian D-criterion for the estimation of the respondent-level effects β_i, which we simply refer to as the "D-criterion" in this paper. Thus, D-optimal designs in this paper strictly refer to the designs for the efficient estimation of the respondent-level effects β_i, with matrices $\{X_i\}$ that maximize

$$D = \int \left\{ \log \left| \sigma^{-2} \mathrm{X}_i^{'} \mathrm{X}_i + \Lambda^{-1} \right| \right\} p(\Lambda) p(\sigma^2) d\Lambda d\sigma^2,$$

Note that matrices $\{Z_i\}$ of contextual variables do not play a part in the identification of D-optimal designs. Furthermore, we demonstrate in Section 4 that efficient designs for the estimation of respondent-level effects may not be efficient for the estimation of hyper-parameters.

In certain situations when all respondents receive the same stimuli (X_i = X), X can be determined independently of $\{Z_i\}$ in an optimal design under the ψ criterion. For example, in a survey study, when the same questionnaire is given to all respondents, the optimal construction of the questionnaire is often independent of the optimal sampling of the respondents on the basis of demographic information such as age or income. However, in other situations when respondents receive customized stimuli or when the construction of the stimuli is linked to the contextual variables, optimal $\{X_i\}$ and $\{Z_i\}$ need to be determined jointly.

In this paper, we focus on the estimation of a function $g(\theta)$ of hyper-parameter vector θ. Using a Taylor series expansion, it follows from the derivation of (7) that, an optimal design for the estimation of $g(\theta)$ is a design with matrices $\{X_i\}$ and $\{Z_i\}$, i=1, …, n, that maximizes

$$\tilde{\psi} = \int \left\{ -\log \left| \left(\frac{\partial g(\theta)}{\partial \theta} \right)^{'} \left(\sum_{i=1}^{n} Z_i^{'} X_i^{'} (\sigma^2 I_{m_i} + X_i \Lambda X_i^{'})^{-1} X_i Z_i \right)^{-1} \left(\frac{\partial g(\theta)}{\partial \theta} \right) \right| \right\} p(\theta) p(\Lambda) p(\sigma^2) d\theta d\Lambda d\sigma^2, \quad (8)$$

where $p(\theta)$ is the probability density function of the prior distribution of θ. The matrix inverse inside the log determinant function is the general inverse $\left(\sum_{i=1}^{n} Z_i^{'} X_i^{'} (\sigma^2 I_{m_i} + X_i \Lambda X_i^{'})^{-1} X_i Z_i \right)^{-}$ for the case that $g(\theta)$ is uniquely identifiable but θ is not.

We use the $\tilde{\psi}$ -criterion to select the survey designs in Section 4 in the context of the level-effect in conjoint analysis where X_i is linked to Z_i. The next section introduces the level-effect, and proposes a hierarchical linear model that captures the level-effect as one of the hyper-parameters.

3. Modeling the Level-effect in Conjoint Analysis

Numerous explanations of the level-effect have been proposed in the literature. Currim, Weinberg and Wittink (1981) find that the ordinal properties of ranking and rating measures contribute to the occurrence of the level-effect. Wittink et al. (1992), Steenkamp and Wittink (1994), Verlegh, Schifferstein and Wittink (2002) find that certain experimental design methods such as the utility-balance approach in ACA (Adaptive Conjoint Analysis), and small sampling errors seem to reduce the magnitude of the level-effect. Steenkamp and Wittink (1994) also investigate the attention-based explanation which states that the respondents' attention to an attribute may increase with the addition of attribute levels. However, only minimal support of the attention-based explanation has been found in empirical studies. Most recently, Verlegh, Schifferstein and Wittink (2002) suggest that the level-effect is most likely due to respondents' tendencies to uniformly distribute their responses over the corresponding continuum of the measurement scale.

In this section, we model the level-effect by incorporating two ideas from the psychology literature – the Range-Frequency theory of Parducci (1965, 1974, 1982) and Krumhansl's (1978) Distance-Density model. We first specify the model for a single attribute, and then express the model in its general form.

The Model: Single-Attribute Case

Parducci (1965) proposes a theory on how evaluation is influenced by two factors – the range and frequency of levels of an attribute. His theory posits that preferences reflect a compromise between these factors, where range is defined in terms of a linear mapping of attribute-levels to a measure of value, and frequency is defined in terms of an ordinal mapping to value. To illustrate, suppose respondents are presented with five possible car prices:

$20,000	$21,000	$26,000	$28,000	$30,000

Evaluations based on the principle of range leads to values of:

10	9	4	2	0

and evaluations based on frequency result in values of:

10	7.5	5	2.5	0

Cooke et al. (2004) formalize Parducci's range-frequency theory into the following model:

$$R_g = \left(S_g - S_{\min}\right) / \left(S_{\max} - S_{\min}\right) \tag{9}$$

$$F_g = \left(Rank\left(g\right) - 1\right) / \left(N - 1\right) \tag{10}$$

$$J_g = \omega R_g + (1-\omega) F_g \qquad (11)$$

where J_g denotes a normalized value of attribute-level g as a weighted average of the range (R_g) and frequency (F_g) effects with $0 \leq \omega \leq 1$. Cooke et al. (2004) transform this normalized value to an actual value by the equation:

$$A_g = b + mJ_g \qquad (12)$$

where b is the respondent's perceived value of the lowest-level of the attribute, and m is the "perceived distance" between the lowest and highest attribute-levels.

We extend the model proposed by Cook et al. (2004), (9) to (12), by modeling *m* according to the Distance-Density model in Krumhansl (1978) which states that "...two points in a relatively dense region of a stimulus space would have a smaller similarity measure than two points of equal interpoint distance but located in a less dense region of the space...". Specifically, we assume *m* is a linear function of the number of attribute levels *L*:

$$m = \kappa + \upsilon L \qquad (13)$$

and the final form of the model is:

$$A_g = b + \kappa F_g + \upsilon L F_g + \kappa\omega\left(R_g - F_g\right) + \upsilon\omega L\left(R_g - F_g\right) + \delta_g \qquad (14)$$

where δ_g is an error term distributed according to a $N(0,\lambda^2)$ distribution that allows for respondent-specific deviation from the assumed structure.

The parameter υ in equation (14) measures the level-effect in conjoint analysis. To estimate υ, it is required that the conjoint study contains multiple parts which have different numbers of attribute levels *L*. The difference $R_g - F_g$ in equation (14) represents how far away the relative range is from the relative rank and provides a measure of the skewness of the distribution of the attribute levels. If an attribute has levels packed on the lower end, for example, (10, 15, 20, 30, 35, 40, 70, 90, 100), then we will observe prevailingly $R_g - F_g < 0$. Similarly, if an attribute has levels that are packed up on the higher end, we will have prevailingly $R_g - F_g > 0$. If an attribute has equally-spaced levels, or when an attribute has only 2 levels, then $R_g - F_g = 0$ for all g. So, for ω in equation (14) to be estimable, we need to have at least one part of the study in which the attribute has more than two levels and the levels are not equally spaced.

The Model: General Expression in Matrix Form

We now express the model in matrix form for the general case of multiple respondents rating multi-attribute products in a conjoint study. The model takes the form of a hierarchical linear model:

$$y_{i,s} \mid \beta_{i,s}, \sigma^2 \sim Normal(\mathrm{X}_{i,s}\beta_{i,s}, \sigma^2 \mathrm{I}), \qquad (15)$$

where s indexes the parts of the conjoint study ($s = 1, 2, \ldots, \mathrm{S}$) with possibly different numbers of attribute levels. $y_{i,s}$ is the vector of profile ratings from respondent i ($i = 1, \ldots, n$) in part s of the study, $\beta_{i,s}$ is the vector of attribute-level part-worths, and $\mathrm{X}_{i,s}$ is the model matrix representing the stimuli presented to respondent i in part s of the study. If all respondents in part s of the study are asked to rate the same set of stimuli, then the model matrix is the same across all respondents in that part of the study, that is, $\mathrm{X}_{i,s} = \mathrm{X}_s$. Note that (15) for a given part s of the study is of the same form as (1), the first level of the hierarchical model introduced in Section 1.

In the second level of the hierarchical model, the part-worths $\beta_{i,s}$ are modeled according to the Range-Frequency/Distance-Density theory as formalized in Equation (14). The expression of the second level of the hierarchical model is:

$$\beta_{i,s} \mid \theta, \Lambda_s \sim Normal\left(\mathrm{Z}_s\theta, \Lambda_s\right), \qquad (16)$$

The hyper-parameter vector θ is defined as

$$\theta' = (\mu, \theta_1', \ldots, \theta_\tau', \ldots, \theta_T'), \qquad (17)$$

where τ indicates the τ^{th} attribute (τ =1, …, T). When attribute τ is monotonic with different numbers of levels across the multiple parts of the study, $\theta_\tau^{'} = (\kappa_\tau, \upsilon_\tau, \kappa_\tau\omega_\tau, \upsilon_\tau\omega_\tau)$, where the set of $(\kappa_\tau, \upsilon_\tau, \omega_\tau)$ corresponds to the $(\kappa, \upsilon, \omega)$ in (14) for the attribute τ. If the monotonic attribute τ has the same number of levels across the multiple parts of the study, υ_τ is not estimable and therefore $\theta_\tau^{'} = (\kappa_\tau, \kappa_\tau\omega_\tau)$. Similarly, if the attribute levels in attribute τ are equally-spaced or only contain the two extreme levels in all parts of the study, ω_τ is not estimable and therefore $\theta_\tau^{'} = (\kappa_\tau, \upsilon_\tau)$. Finally, for non-monotonic attributes such as brand name or color, the vector θ_τ simply contains the means of the individual-level-effects over the respondents.

The covariate matrix Z_s in (16) is of a block diagonal structure:

$$Z_s = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & Z_{1,s} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & Z_{\tau,s} & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & Z_{T,s} \end{pmatrix} \quad (18)$$

When attribute τ is a monotonic with different numbers of levels across the multiple parts of the study, $Z_{\tau,s} = 1$. $Z_{\tau,s}$ is a matrix that contains of F, $L_{\tau,s}F$, $(R-F)$ and $L_{\tau,s}(R-F)$ on the various levels of attribute τ, as required by Equation (14),. $Z_{\tau,s}$ contains only F and $R-F$ on the various levels of attribute τ if the monotonic attribute τ has the same number of levels across studies, and $Z_{\tau,s}$ consists only of F on various levels of the attribute τ if the attribute levels in attribute τ are equally-spaced or only contain the two extreme levels in all parts of the study.

Appendix A illustrates the model matrices $X_{i,s}$ and covariate matrices Z_s in our hierarchical model (15) and (16) for an investigation of the level-effect of credit card interest rate (APR). The designs were selected using the $\tilde{\psi}$ -criterion in equation (8) that accounts for the dependence of Z_i

on $X_{i,s}$ due to the level-effect. Next we illustrate the importance of accounting for this dependence when estimating contextual factors that influence consumer valuation of product offerings.

4. Survey Designs for Studying the Level-effect

A survey study was conducted through Harris Interactive, a global marketing research firm, to evaluate the effectiveness of the proposed model and design criterion. Credit card products were used as stimuli in the survey. Measurement of the level-effect in credit-card preferences requires that the study contains multiple parts where the number of attribute levels varies. The attributes used in the survey studies were the APR (interest rate) of the card, the card provider and the reward program. Part 1 of the study involved the two extreme levels (8.99% and 17.99%) of APR in the study, part 2 involved four levels of APR with the addition of two intermediate levels, and part 3 involved three levels of APR with the addition of one intermediate level. The numbers of levels of the other two attributes, card provider and reward program, remained the same across the three parts of the study. In particular, there were two card providers -- Capital One vs. Citibank, and three levels of rewards – none, cash reward and travel reward. Data from parts 1 and 2 of the study were used to estimate the model parameters, and part 3 was reserved for holdout predictive testing. Each respondent in the survey was asked to evaluate 12 credit card products on a 0 to 10 rating scale, with 0 being least likely to apply and 10 being most likely to apply for the card.

The vector θ of the hierarchical model in (17) is

$$\theta = (\mu, \kappa_A, \upsilon_A, \kappa_A\omega_A, \upsilon_A\omega_A, \theta_P, \theta_C, \theta_T)', \tag{19}$$

where the $(\kappa_A, \upsilon_A, \omega_A)$ reflect the $(\kappa, \upsilon, \omega)$ in (14) on the APR attribute, θ_P is the mean contrast between the two card providers, θ_C is the mean contrast between cash reward and no reward, and θ_T is the mean contrast between travel reward and no reward. The vector θ is a nonlinear function of θ^* which contains the unique parameters to be estimated in (19), that is,

$$\theta^* = (\mu, \kappa_A, \upsilon_A, \omega_A, \theta_P, \theta_C, \theta_T)'. \tag{20}$$

For efficient estimation of θ^* through parts 1 and 2 of the study, the construction of the stimuli (which determines $\{X_{i,s}\}$) and the physical values of APR (which determines $\{Z_s\}$) for the two parts need to be jointly determined so that $\tilde{\psi}$ in (8) is maximized. Note that Z_1 in part 1 of the study was fixed since the two extreme APR levels were fixed at 8.99% and 17.99%. To simplify the problem, the same set of stimuli was presented to the respondents in the same part of the study ($X_{i,s}=X_s$), and a full-factorial design with twelve profiles, as shown in Table A.1 of Appendix A, was used in part 1 of the survey. Therefore, X_1 was fixed, and the problem was simplified to the search of X_2 and Z_2 for part 2 of the study that maximize

$$\tilde{\psi} = \int \left\{ -\log \left| \left(\frac{\partial g(\theta)}{\partial \theta} \right)' \left(\sum_{s=1}^{2} Z_s' X_s' (\sigma_s^2 I_{12} + X_s \Lambda_s X_s')^{-1} X_s Z_s \right)^{-} \left(\frac{\partial g(\theta)}{\partial \theta} \right) \right| \right\}$$

$$\times p(\theta) p(\Lambda_1) p(\Lambda_2) p(\sigma_1^2) p(\sigma_2^2) d\theta d\Lambda_1 d\Lambda_2 d\sigma_1^2 d\sigma_2^2, \qquad (21)$$

given an equal number of respondents in each part of the study, where

$$\frac{\partial g(\theta)}{\partial \theta} = \frac{\partial \theta^*}{\partial \theta} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1/\omega_A & 0 & 1/\kappa_A & 0 & 0 & 0 \\ 0 & 0 & 1/\omega_A & 1/\nu_A & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad (22)$$

Based on the credit card offers available on the market, there were five possible choices of intermediate APR levels to be included in the study -- 9.99%, 11.99%, 12.99%, 14.99%, and 15.99%. Therefore, for part 2 of the study, there were a total of $\binom{5}{2} = 10$ possible choices of Z_2. For each possible choice of Z_2, we obtained an optimal X_2 that maximizes $\tilde{\psi}$ in (21) through computer search as explained below. Then we recorded the $\tilde{\psi}$ value for each pair of (X_2, Z_2), and compared the $\tilde{\psi}$ values to obtain the optimal pair of X_2 and Z_2. Similarly, for part 3 of the study, there were a total of 5 possible choices of Z_3. We used the same approach to obtaining the optimal X_3 and Z_3 by replacing the index $s = 2$ with the index $s = 3$ in (21).

Table 1 lists the attributes and attribute-levels in the three parts of the study.

Part 1	APR:	(8.99%, 17.99%)
	Provider:	(Capital One, Citibank)
	Reward:	(None, Cash, Travel)
Part 2	APR:	(8.99%, 9.99%, 15.99%, 17.99%)
	Provider:	(Capital One, Citibank)
	Reward:	(None, Cash, Travel)
Part 3	APR:	(8.99%, 9.99%, 17.99%)
(Holdout)	Provider:	(Capital One, Citibank)
	Reward:	(None, Cash, Travel)

Table 1: Web-based 12-Profile Credit Card Survey Study

Table A.6 in Appendix A reports the matrices Z_2 and Z_3 in the optimal pairs (X_2, Z_2) and (X_3, Z_3) that were selected for parts 2 and 3 of the study, together with the fixed Z_1 for part 1 of the study.

Simple exchange algorithms (see Atkinson and Donev, 1992) have been used in the literature for the search of D-optimal designs. We modified a simple exchange algorithm to obtain an optimal X_2 that maximizes $\tilde{\psi}$ in (21) for given X_1, Z_1 and Z_2, where the integral of $\tilde{\psi}$ was calculated using Monte Carlo method. Based on prior knowledge, under the standardized orthogonal effects coding (see Kuhfeld, 2005) of the model matrices X_1 and X_2, the variance-covariance matrices Λ_1 and Λ_2 were expected to have positive off-diagonal elements, that is, the random effects were expected to be positively correlated although the actual sizes of the correlations and the variances were unknown. Following recommendations in Liu et al. (2007), the following priors were used, respectively, for Λ_1 and Λ_2, in the computer search of an optimal X_2.

$$\Lambda_1 \sim \text{Inverted Wishart}(\nu_0{=}7,\ V_0{=}2I_5 + J_5),$$

$$\Lambda_2 \sim \text{Inverted Wishart}(\nu_0{=}9,\ V_0{=}2I_7 + J_7),$$

where J denotes a matrix with all elements being ones. We note that, while these values of V_0 were used in the search of an optimal X_2, alternative values lead to the same rank ordering of candidate X_2 matrices for a wide range of V_0 values where $V_0 = aI + bJ$ with any positive a and b such that $0.5 < a/b \leq 2$. For the rest of the parameters, non-informative priors were used as follows,

$$\sigma_1^2, \sigma_2^2 \sim \text{Inverse Gamma } (3/2, 1/2),$$

$$\omega_A \sim \text{Uniform } (0, 1),$$

$$\theta^*_{-\omega_A} \sim \text{Normal}(0, 100I_6),$$

where $\theta^*_{-\omega_A} = (\mu, \kappa_A, \nu_A, \theta_P, \theta_C, \theta_T)'$.

For purpose of comparison, in addition to an optimal X_2, three other versions of designs with high efficiency for the stimuli construction in part 2 of the study were selected from the same computer search algorithm using different starting points. Tables A.2 and A.3 in Appendix A list the four versions of designs and the corresponding model matrices, where Design IV denotes the optimal design and Design I, II, III denote the three relatively less efficient designs obtained through computer search. Using Design I as the baseline design, a relative $\tilde{\psi}$-efficiency is reported for a design with model matrix $X_2^{\#}$, as defined by

$$\text{Relative } \tilde{\psi}\text{-eff.} = \exp\{\tilde{\psi}(X_2^{\#}) - \tilde{\psi}(X_2^{b})\}, \tag{23}$$

where X_2^{b} denotes the model matrix corresponding to Design I, and $\tilde{\psi}(X_2^{\#})$ and $\tilde{\psi}(X_2^{b})$ represent the values of $\tilde{\psi}$ in (21) corresponding to the design with model matrix $X_2^{\#}$ and X_2^{b}, respectively. As a comparison, relative D-efficiency between the two designs is also reported, as defined by

$$\text{Relative D-eff.} = \exp\left\{\int \left[\log\left|\sigma^{-2}X_2^{\#\prime}X_2^{\#} + \Lambda^{-1}\right| - \log\left|\sigma^{-2}X_2^{b\prime}X_2^{b} + \Lambda^{-1}\right|\right] p(\Lambda)p(\sigma^2)\,d\Lambda\, d\sigma^2\right\}, \tag{24}$$

which measures the relative efficiency of the two designs for the estimation of the respondent-level effects β_i.

As shown in Table 2, designs efficient for the estimation of respondent-level effects (D-efficient) may not be efficient for the estimation of hyper-parameters ($\tilde{\psi}$-efficient). Specifically, Design IV is relatively the most $\tilde{\psi}$-efficient and the least D-efficient, while Design II is relatively the least $\tilde{\psi}$-efficient and the most D-efficient. Design I and Design III are in the middle.

Relative Efficiency	Design I	Design II	Design III	Design IV
$\tilde{\psi}$	100%	95.5%	98.4%	100.4%
D	100%	105.1%	101.4%	93.1%

Table 2: Relative Efficiencies of Designs for Stimuli Construction in Part 2 of the Study

Similarly, three versions of designs were selected for stimuli construction in part 3 of the study, as reported in Appendix A. Part 3 of the study was used as holdout for prediction validation. Namely, data obtained from parts 1 and 2 of the study were used to estimate the θ^* in (20), and the θ^* estimates were used to predict the mean preferences (over the respondents) in the holdout study (see Section 5).

In general, $\tilde{\psi}$-efficient designs differ from D-efficient designs in this context in three ways. First, standard D-efficient designs tend to be orthogonal, while $\tilde{\psi}$-efficient designs (for stimuli specification X) are only orthogonal for independent random-effects and fixed contexts Z which are independent of X, under non-informative priors. Departures from any of these conditions can result in non-orthogonal optimal designs. When ω equals one, such that respondents depend solely on the range of the attribute levels, equally spaced attribute levels tend to provide a more efficient design. When ω equals zero, such that respondents depend solely on the relative ranks of the attribute levels, the values of the attribute levels do not matter.

5. Empirical Results

The data were collected via a national web-based survey through Harris Interactive. The following screening rules were used during the data collection: a qualified respondent needs to be 18 years of age or older, live in U.S., and have a credit card issued under his/her own name. A total of 1000 respondents participated in the study. Respondents were randomly assigned to the different parts of the study, and the presentation order of the 12 credit card profiles was also random. The data showed that some respondents gave the same ratings to all the 12 profiles presented to them, some respondents chose one product and rated the remaining eleven products 0, and some respondents used the scale in a reverse order (i.e., 0 was used for most likely to apply for the card, and 10 was used for least likely to apply). Responses from these respondents were considered invalid and taken out of the data. The final data contained 757 valid respondents, averaging about 95 valid respondents per design version of the survey.

An MCMC method was used to obtain the estimates of θ^* in (20) where a Metropolis-Hastings (M-H) algorithm was used within a Gibbs sampler to generate the posterior draws. Appendix B provides the details of the algorithm. Using the responses from part 1 and from each design version of part 2 of the study, the proposed hierarchical model was fit to the data and the posterior estimates of θ^* were obtained, as reported in Table 3 with the posterior means and the posterior standard deviations in parenthesis.

θ^*	Design I	Design II	Design III	Design IV
μ	0.325 (0.117)	0.234 (0.115)	0.254 (0.112)	0.202 (0.112)
$\varkappa_A$	3.427 (0.499)	3.270 (0.512)	3.495 (0.506)	3.174 (0.520)
ν_A	0.224 (0.138)	0.289 (0.150)	0.240 (0.141)	0.373 (0.143)
ω_A	0.925 (0.070)	0.891 (0.092)	0.922 (0.070)	0.943 (0.054)
θ_P	-0.048 (0.123)	0.061 (0.125)	-0.015 (0.129)	-0.126 (0.120)
θ_C	1.595 (0.180)	1.705 (0.175)	1.414 (0.164)	1.684 (0.172)
θ_T	0.529 (0.163)	0.508 (0.163)	0.681 (0.151)	0.631 (0.153)

Table 3: Posterior Estimates of θ^* From Part 1 and Each Design Version of Part 2 of the Study

In all four design versions, the Bayesian 95% highest probability density (HPD) regions of ν_A (the level-effect of APR) do not include 0, confirming the existence of the level-effect. That is, the contrast between the 8.99% and the 17.99% APR increases with the addition of the two intermediate APR levels. Estimation results on the other parameters are quite consistent in all occasions. For example, estimates on credit card provider (θ_P) suggest that respondents do not have a preference of one provider (Capital One) over another (Citibank); estimates on rewards (θ_C and θ_T) suggest that respondents prefer cash reward over none and travel over none.

The parameter estimates reported in Table 3 indicate that the level-effect parameter (ν_A) is estimated to have a posterior mean of 0.28 on average, and that the range-frequency weight (ω_A) in (20) has a posterior mean of 0.92 on average. This implies that attribute-levels in this study map onto the range-frequency value J_g mostly through the relative range R_g. Also, for each addition of an intermediate APR level, the mean perceived distance between the highest and lowest APR levels increases by 0.28. That is, the mean contrast between the highest and lowest APR in part 2 of the study is 0.28 × 2 = 0.56 units larger than in part 1 of the study due to the addition of two intermediate APR levels. Thus, it is estimated to be large relative to the other effect sizes -- approximately as large as the mean contrast between a travel reward and no reward.

Although earlier empirical studies on category ratings found the range-frequency weight (ω_A) close to 0.5 (Wedell et al. 1989, page 235), Parducci and Wedell (1986) investigated situations that influence the magnitude of this parameter. Results from their experiments showed that the weight can sometimes be as high as 91% (Table 3, Parducci and Wedell, 1986), suggesting a range-driven behavior rather than frequency-driven. They suggest that the range-frequency weight is affected by the difficulty of recognizing frequency values by the respondents. This occurs when respondents are asked to rate on 20- or 100-point scales, rather than on 5- or 7-point scales, or when the profiles under evaluation contain multiple attributes each with several levels as in conjoint analysis.

Accuracy of Estimates

We compare the efficacy of the four different designs using two measures of accuracy -- the determinant of the variance-covariance matrix (DETVAR) of the posterior estimates of θ^* normalized by the dimension of θ^*, and the average variance (AVGVAR) of the posterior estimates of θ^* normalized by the size of θ^*. Results are displayed in Table 4. We note that without knowledge of the actual values of θ^*, these two measures provide direct measures of the estimation efficiency by assuming that the parameter estimates are unbiased.

Both the DETVAR and the AVGVAR measures show consistency with the relative $\tilde{\psi}$-efficiencies of the designs, that is, designs with higher $\tilde{\psi}$-efficiency lead to smaller variances of the θ^* estimates as measured by DETVAR and AVGVAR. Table 4 also shows that neither of the performance measures is consistent with the relative D-efficiencies of the designs. Note especially designs II and IV, where design IV is relatively the most $\tilde{\psi}$-efficient and design II is the most D-efficient among the four designs. The performance measures give evidence for the use of the $\tilde{\psi}$ criterion in designing experiments for efficient estimation of hyper-parameters, rather than the use of the D criterion.

Performance Measure	Design I Rel. $\tilde{\psi}$ =100% Rel. D=100%	Design II Rel. $\tilde{\psi}$ =95.5% Rel. D=105.1%	Design III Rel. $\tilde{\psi}$ =98.4% Rel. D=101.4%	Design IV Rel. $\tilde{\psi}$ =100.4% Rel. D =93.1%
DETVAR Relative DETVAR Efficiency	0.013 100%	0.017 76.5%	0.014 92.9%	0.013 100%
AVGVAR Relative AVGVAR Efficiency	0.046 100%	0.054 85.2%	0.051 90.2%	0.046 100%

Table 4: Performance Comparison of the Four Designs in Part 2 of the Study

In-Sample Fit

We compare the fit of the proposed model to that of a number of alternative models. Model 1 estimates parameters from parts 1 and 2 of the data independently and does not attempt to model the level-effect:

Model 1: $$y_{i,s} \sim N\left(X_{i,s}\beta_{i,s}, \tau_s\right), \ \ \beta_{i,s} \sim N\left(\bar{\beta}_s, \Lambda_s\right) \tag{25}$$

In model 2, the APR attribute is treated as a continuous variable without regard to the presence of the level-effect:

Model 2: $$y_i \sim N\left(X_i\beta_i, \tau\right), \ \ \beta_i \sim N\left(\bar{\beta}, \Lambda\right) \tag{26}$$

Model 3 retains the dummy-variable coding for APR, but treats the APR levels as a continuous variable in the hyper-parameter specification, which is essentially the same as setting the range-frequency weight ω_A= 1 in our model, i.e.,

Model 3: $$y_{i,s} \sim N\left(X_{i,s}\beta_{i,s}, \tau_s\right), \ \ \beta_{i,s} \sim N\left(Z\theta, \Lambda\right) \tag{27}$$

Models 2 and 3 allow for projection to new contexts, while model 1 does not. In-sample fit statistics are reported in table 5, and indicate that our proposed model fits the data best in terms of log marginal density. Moreover, treating APR as a continuous variable in either the lower-level (Model 2) or the upper-level (Model 3) of the hierarchical model substantially degrade the fit, indicating that contextual effects are present.

Model for Comparison	Design I Rel. $\tilde{\psi}$ =100% Rel. D=100%	Design II Rel. $\tilde{\psi}$ =95.5% Rel. D=105.1%	Design III Rel. $\tilde{\psi}$ =98.4% Rel. D=101.4%	Design IV Rel. $\tilde{\psi}$ =100.4% Rel. D =93.1%
Proposed Model	-4515	-4455	-4414	-4258
Model 1:	-4526	-4491	-4438	-4296
Model 2:	-4583	-4803	-4446	-4615
Model 3	-4600	-4744	-4483	-4559

Table 5: In-Sample Fit Based on Log Marginal Density

Predictive Performance

Data from part 3 of the study are used as validation data to check how well the model predicts to a new context with a different number of APR levels. Estimates of θ^* obtained in Table 3 from each design k (k = 1,…,4) of part 2 of the study were used to predict the mean ratings of the 12 profiles in each design version j (j = 1, 2, 3) of part 3 of the study:

$$\hat{\bar{y}}_{3jk} = \int X_{3j} Z_3 \theta \times \pi(\theta, \Lambda_1, \Lambda_2, \sigma_1^2, \sigma_2^2 \mid \{y_{1i}, y_{2ik}\}, X_1, X_{2k}, Z_1, Z_2) d\theta d\Lambda_1 d\Lambda_2 d\sigma_1^2 d\sigma_2^2$$

where X_{3j} is the model matrix in design version j (j = 1, 2, 3) of part 3 of the study, as shown in Table A.5 of Appendix A. X_1 is the model matrix in part 1, and X_{2k} is the model matrix in design k (k = 1, …, 4) of part 2 of the study, as shown respectively in Table A.1 and A.3 of Appendix A. y_{1i} is the vector of responses from respondent i in part 1, and y_{2ik} is the vector of responses from respondent i in design k of part 2 of the study. Appendix C provides the prediction results on the three design versions of part 3 of the study. For most profiles, the actual mean ratings were within the 95% Bayesian highest posterior density region of the predictions. Prediction accuracy is measured with expected squared error loss (MSE):

$$MSE_{jk} = \int (\bar{y}_{3j} - X_{3j} Z_3 \theta)' (\bar{y}_{3j} - X_{3j} Z_3 \theta)$$

$$\times \pi(\theta, \Lambda_1, \Lambda_2, \sigma_1^2, \sigma_2^2 \mid \{y_{1i}, y_{2ik}\}, X_1, X_{2k}, Z_1, Z_2) d\theta d\Lambda_1 d\Lambda_2 d\sigma_1^2 d\sigma_2^2 \quad (28)$$

where $\bar{y}_{3j}$ is the actual mean ratings of the 12 profiles in design version j (j = 1, 2, 3) of part 3 of the study.

Table 6 displays predictive fits using (28) and parameter values from our proposed model and models 2 and 3 (Expressions 26 and 27). We find that, on average, our proposed model predicts most accurately. The average predictive MSE across all designs is essentially the same as that reported in table 6 for Design IV, the design with highest relative efficiency.

Model for Comparison	Design I Rel. $\tilde{\psi}$ =100% Rel. D=100%	Design II Rel. $\tilde{\psi}$ =95.5% Rel. D=105.1%	Design III Rel. $\tilde{\psi}$ =98.4% Rel. D=101.4%	Design IV Rel. $\tilde{\psi}$ =100.4% Rel. D =93.1%
Proposed Model	0.198	0.209	0.161	0.189
Model 2:	0.199	0.201	0.183	0.194
Model 3	0.200	0.198	0.180	0.194

Table 6: MSE of Mean Rating Predictions in the Holdout Part (Part 3) of the Study

6. Discussion and Concluding Remarks

The level-effect is an example of a broad class of problems in Marketing and other disciplines that involve the learning of effect sizes and what drives big or small effects. Learning of the level-effect gives us ideas on how consumer sensitivity to a product attribute varies with the number of attribute levels, and allows us to predict to a new context. We propose in this paper a hierarchical linear model that models individual consumer behavior and, by incorporating a number of ideas from the Psychology literature, models the level-effect as a hyper-parameter.

We designed a credit card survey study in which the survey designs were selected according to the $\tilde{\psi}$ -criterion introduced in (8) for efficient estimation of hyper-parameters. The discussion and choice of design criteria for estimation of hyper-parameters is an under-developed topic in marketing research and, with the advent of modern Bayesian statistical methods, the estimation of models characterizing the variation of effect sizes across contexts and environments is likely to become more prevalent. Our design criterion accommodates interactions between individual-level variables and contextual variables as reflected through matrices X_i and Z_i in hierarchical models. The interactions are present when studying contextual factors that influence consumer preference and choice. We demonstrate that designs efficient under our criterion lead to more accurate hyper-parameter estimation, and that designs efficient for the estimation of respondent-level effects may not be efficient for the estimation of hyper-parameters.

From our survey data, we estimate a large contextual level-effect that points to increased sensitivity to attribute-levels as the number of levels increases. The level-effect has implications for merchandising because it implies that consumer preference for product features is dependent on the assortment of choice. Our results imply that a high-assortment retailer will find that consumers express greater preference among attribute levels, while a low-assortment retailer will find that preferences are not as well defined and that consumers are more willing to switch among choice alternatives. This finding is consistent with the merchandising policies of discount retailers, such as WalMart, who typically offer low levels of assortment. We note that the level-effect we examined is on a monotonic attribute, not one for which an ideal-point preference exists. We have not examined if the level-effect exists for nominal attributes such as the color – e.g., that increasing choices of colors on cars makes people more sensitive to car color when making a purchase decision.

Hierarchical models offer a convenient structure for pooling results across studies to quantify changes in effect-sizes. In the past, these changes have typically not been associated with covariates, Z_i, that describe aspects of the study from which the data originate. For example, while researchers routinely compare responses for alternative stimuli, the comparison typically involves examination of simple contrasts, or, equivalently, assumes that Z_i in the hierarchical model is dummy-variable coded. Hierarchical models allow for a more general coding scheme that enables the study of covariates that are more generally associated with experiment conditions, as illustrated in our empirical study. The design criterion presented here provides a means of evaluating designs for efficient learning about these contextual effects.

In this paper, we focused on a design criterion for efficient hyper-parameter estimation. An interesting direction for future research is to investigate optimal designs under alternative design criteria that may be of interest. For example, if all possible new contexts are known beforehand, then a good criterion to use is a criterion for the predictions, such as one that minimizes the average variance of the predictions. Additional topics for future research include the study of incomplete block designs in which different respondents receive different sets of stimuli. Furthermore, a natural extension of the current research is to the choice designs in the setting of hierarchical non-linear models.

APPENDIX A: Survey Designs and Corresponding Model Matrices

APR	Provider	Rewards	Corresponding Model Matrix X_1				
17.99%	Capital One	Travel	1	0	1	0	1
17.99%	Capital One	Cash	1	0	1	1	0
17.99%	Capital One	None	1	0	1	0	0
17.99%	Citibank	Travel	1	0	0	0	1
17.99%	Citibank	Cash	1	0	0	1	0
17.99%	Citibank	None	1	0	0	0	0
8.99%	Capital One	Travel	1	1	1	0	1
8.99%	Capital One	Cash	1	1	1	1	0
8.99%	Capital One	None	1	1	1	0	0
8.99%	Citibank	Travel	1	1	0	0	1
8.99%	Citibank	Cash	1	1	0	1	0
8.99%	Citibank	None	1	1	0	0	0

Table A.1: Full-factorial design for stimuli construction in part 1 of the study and corresponding model matrix under dummy coding

DESIGN I

APR	Provider	Rewards
17.99%	Capital One	Cash
17.99%	Citibank	Travel
17.99%	Citibank	None
15.99%	Capital One	Travel
15.99%	Capital One	None
15.99%	Citibank	Cash
9.99%	Capital One	Travel
9.99%	Capital One	Cash
9.99%	Citibank	None
8.99%	Capital One	None
8.99%	Citibank	Travel
8.99%	Citibank	Cash

DESIGN II

APR	Provider	Rewards
17.99%	Capital One	Travel
17.99%	Capital One	Cash
17.99%	Citibank	None
15.99%	Capital One	None
15.99%	Citibank	Travel
15.99%	Citibank	Cash
9.99%	Capital One	Travel
9.99%	Capital One	Cash
9.99%	Citibank	None
8.99%	Capital One	None
8.99%	Citibank	Travel
8.99%	Citibank	Cash

DESIGN III

APR	Provider	Rewards
17.99%	Capital One	Travel
17.99%	Citibank	Cash
17.99%	Citibank	None
15.99%	Capital One	Cash
15.99%	Capital One	None
15.99%	Citibank	Travel
9.99%	Capital One	Cash
9.99%	Citibank	Travel
9.99%	Citibank	None
8.99%	Capital One	Travel
8.99%	Capital One	None
8.99%	Citibank	Cash

DESIGN IV

APR	Provider	Rewards
17.99%	Capital One	Cash
17.99%	Citibank	Travel
17.99%	Citibank	None
15.99%	Capital One	Travel
15.99%	Capital One	None
15.99%	Citibank	Cash
9.99%	Capital One	Travel
9.99%	Capital One	Cash
9.99%	Citibank	None
8.99%	Capital One	Cash
8.99%	Capital One	None
8.99%	Citibank	Travel

Table A.2: Four versions of designs for stimuli construction in part 2 of the study

Design I

1	0	0	0	1	1	0
1	0	0	0	0	0	1
1	0	0	0	0	0	0
1	1	0	0	1	0	1
1	1	0	0	1	0	0
1	1	0	0	0	1	0
1	0	1	0	1	0	1
1	0	1	0	1	1	0
1	0	1	0	0	0	0
1	0	0	1	1	0	0
1	0	0	1	0	0	1
1	0	0	1	0	1	0

Design II

1	0	0	0	1	0	1
1	0	0	0	1	1	0
1	0	0	0	0	0	0
1	1	0	0	1	0	0
1	1	0	0	0	0	1
1	1	0	0	0	1	0
1	0	1	0	1	0	1
1	0	1	0	1	1	0
1	0	1	0	0	0	0
1	0	0	1	1	0	0
1	0	0	1	0	0	1
1	0	0	1	0	1	0

Design III

1	0	0	0	1	0	1
1	0	0	0	0	1	0
1	0	0	0	0	0	0
1	1	0	0	1	1	0
1	1	0	0	1	0	0
1	1	0	0	0	0	1
1	0	1	0	1	1	0
1	0	1	0	0	0	1
1	0	1	0	0	0	0
1	0	0	1	1	0	1
1	0	0	1	1	0	0
1	0	0	1	0	1	0

Design IV

1	0	0	0	1	1	0
1	0	0	0	0	0	1
1	0	0	0	0	0	0
1	1	0	0	1	0	1
1	1	0	0	1	0	0
1	1	0	0	0	1	0
1	0	1	0	1	0	1
1	0	1	0	1	1	0
1	0	1	0	0	0	0
1	0	0	1	1	1	0
1	0	0	1	1	0	0
1	0	0	1	0	0	1

Table A.3: Corresponding model matrix (X_2, dummy-coded) in part 2 of the study

DESIGN I			DESIGN II			DESIGN III		
APR	**Provider**	**Rewards**	**APR**	**Provider**	**Rewards**	**APR**	**Provider**	**Rewards**
17.99%	CapOne	Cash	17.99%	CapOne	Travel	17.99%	CapOne	Cash
17.99%	CapOne	None	17.99%	CapOne	None	17.99%	CapOne	None
17.99%	Citibank	Travel	17.99%	Citibank	Cash	17.99%	Citibank	Travel
17.99%	Citibank	Cash	17.99%	Citibank	None	17.99%	Citibank	Cash
9.99%	CapOne	Travel	9.99%	CapOne	Cash	17.99%	Citibank	None
9.99%	CapOne	None	9.99%	CapOne	None	9.99%	CapOne	Travel
9.99%	Citibank	Cash	9.99%	Citibank	Travel	9.99%	CapOne	None
9.99%	Citibank	None	9.99%	Citibank	Cash	9.99%	Citibank	Cash
8.99%	CapOne	Travel	8.99%	CapOne	Travel	8.99%	CapOne	Travel
8.99%	CapOne	Cash	8.99%	CapOne	Cash	8.99%	CapOne	Cash
8.99%	Citibank	Travel	8.99%	Citibank	Travel	8.99%	Citibank	Travel
8.99%	Citibank	None	8.99%	Citibank	None	8.99%	Citibank	None

Table A.4: Three versions of designs for stimuli construction in part 3 of the study

Design I

1	0	0	1	1	0
1	0	0	1	0	0
1	0	0	0	0	1
1	0	0	0	1	0
1	1	0	1	0	1
1	1	0	1	0	0
1	1	0	0	1	0
1	1	0	0	0	0
1	0	1	1	0	1
1	0	1	1	1	0
1	0	1	0	0	1
1	0	1	0	0	0

Design II

1	0	0	1	0	1
1	0	0	1	0	0
1	0	0	0	1	0
1	0	0	0	0	0
1	1	0	1	1	0
1	1	0	1	0	0
1	1	0	0	0	1
1	1	0	0	1	0
1	0	1	1	0	1
1	0	1	1	1	0
1	0	1	0	0	1
1	0	1	0	0	0

Design III

1	0	0	1	1	0
1	0	0	1	0	0
1	0	0	0	0	1
1	0	0	0	1	0
1	0	0	0	0	0
1	1	0	1	0	1
1	1	0	1	0	0
1	1	0	0	1	0
1	0	1	1	0	1
1	0	1	1	1	0
1	0	1	0	0	1
1	0	1	0	0	0

Table A.5: Corresponding model matrix (X_3, dummy-coded) in part 3 of the study

Study 1 Covariate Matrix Z_1

1	0	0	0	0	0	0	0
0	1	2	0	0	0	0	0
0	0	0	0	0	1	0	0
0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	1

Study 3 Covariate Matrix Z_3

1	0	0	0	0	0	0	0
0	0.5	1.5	0.39	1.17	0	0	0
0	1	3	0	0	0	0	0
0	0	0	0	0	1	0	0
0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	1

Study 2 Covariate Matrix Z_2

1	0	0	0	0	0	0	0
0	0.33	1.33	-0.11	-0.44	0	0	0
0	0.67	2.67	0.22	0.88	0	0	0
0	1	4	0	0	0	0	0
0	0	0	0	0	1	0	0
0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	1

Table A.6: Covariate matrices for the three parts of the study

APPENDIX B: MCMC Algorithm for Model Fitting

The following standard prior distribution assumptions (see, for example, Gamerman, 1997; Rossi, Allenby and McCulloch, 2005) are made on parameters ω_A, $\theta^*_{-\omega_A} = (\mu, \kappa_A, \nu_A, \theta_P, \theta_C, \theta_T)'$, $\{\Lambda_s\}$ and $\{\sigma_s^2\}$ $(s = 1, 2)$:

$$\omega_A \sim \text{Uniform } (0, 1)$$

$$\theta^*_{-\omega_A} \sim \text{Normal}\,(b_0 = 0, D_0^{-1} = 100\mathrm{I}_6)$$

$$\Lambda_s \sim \text{Inverted Wishart}\,(\eta_{0,s} = p_s + 3, \Delta_{0,s} = \eta_{0,s}\mathrm{I}_{p_s})$$

$$\sigma_s^2 \sim \text{Inverse Gamma}\,(\frac{w_0}{2}, \frac{W_{0,s}}{2})$$

Where $p_1 = 5$, $p_2 = 7$, $w_0 = 3$, and $W_{0,s} = w_0 * \sum_{i=1}^{n}\sum_{j=1}^{m_i}(y_{ij} - \bar{y}_{..})^2 / (m_s n_s - 1)$. Here, m_s represents the number of profiles and n_s represents the number of respondents in the study s (s=1,2). In the 12-profile credit card survey studies 1 and 2, $m_1 = m_2 = 12$. A Metropolis-Hastings (M-H) algorithm is used within a Gibbs sampler. The estimation algorithm proceeds by recursively generating draws from the following densities:

1. Independently generate $\{\beta_{i,s}$, $i = 1, \ldots, n$, $s = 1, 2\}$ from the following multivariate normal distribution:

$$[\beta_{i,s} \mid y_{i,s}, X_{i,s}, Z_s, \theta, \Lambda_s, \sigma_s^2] \propto [y_{i,s} \mid X_{i,s}, \beta_{i,s}, \sigma_s^2][\beta_{i,s} \mid Z_s, \theta, \Lambda_s] \sim \text{Normal }(b_{i,s}, D_{i,s}),$$

$$D_{i,s} = (\sigma_s^{-2} X'_{i,s} X_{i,s} + \Lambda_s^{-1})^{-1}, \qquad b_{i,s} = D_{i,s}(\sigma_s^{-2} X'_{i,s} y_{i,s} + \Lambda_s^{-1} Z_s \theta).$$

2. Generate $\{\sigma_s^2$, $s = 1, 2\}$ from the following Inverted chi-squared distribution

$$[\sigma_s^2 \mid \{y_{i,s}, X_{i,s}, \beta_{i,s}\}] \propto [\{y_{i,s}\} \mid \{X_{i,s}, \beta_{i,s}, \sigma_s^2\}][\sigma_s^2] \sim \text{Inverted Chi-squared } (w_s, W_s)$$

$$w_s = w_{0,s} + m_s n_s , \quad W_s^{-1} = W_{0,s}^{-1} + \sum_{i=1}^{n_s} (y_{i,s} - X_{i,s}\beta_{i,s})' (y_{i,s} - X_{i,s}\beta_{i,s}) .$$

3. Generate $\theta^*_{-\omega_A}$

$$[\theta^*_{-\omega_A} \mid \{Z_s, \beta_{i,s}, \Lambda_s\}, \omega_A] \propto [\{\beta_{i,s}\} \mid \{Z_s, \Lambda_s\}, \theta][\theta^*_{-\omega_A}],$$

where $[\{\beta_{i,s}\} \mid \{Z_s, \Lambda_s\}, \theta]$ is the product of multivariate normal densities, $\prod_{s=1}^{S} \prod_{i=1}^{n} [\beta_{i,s} \mid Z_s, \theta, \Lambda_s]$.

A random-walk M-H is used to generate the draws of $\theta^*_{-\omega_A}$. Let

$$\theta^{*(k)}_{-\omega_A} = (\mu^{(k)}, \kappa_A^{(k)}, \nu_A^{(k)}, \theta_P^{(k)}, \theta_C^{(k)}, \theta_T^{(k)})'$$

be the *k*th draw; the next draw is given by

$$\theta^{*(k+1)}_{-\omega_A} = \theta^{*(k)}_{-\omega_A} + \delta,$$

where δ is a draw from the candidate generating density Normal(0,0.05^{2}I). Let

$$\theta^{(k)} = (\mu^{(k)}, \kappa_A^{(k)}, \nu_A^{(k)}, \kappa_A^{(k)}\omega_A, \nu_A^{(k)}\omega_A, \theta_P^{(k)}, \theta_C^{(k)}, \theta_T^{(k)})' .$$

The probability of accepting the new draw $\theta^{*(k+1)}_{-\omega_A}$ is given by

$$\min\left[\frac{[\{\beta_{i,s}\} \mid \{Z_s, \Lambda_s\}, \theta^{(k+1)}][\theta^{*(k+1)}_{-\omega_A}]}{[\{\beta_{i,s}\} \mid \{Z_s, \Lambda_s\}, \theta^{(k)}][\theta^{*(k)}_{-\omega_A}]}, \quad 1 \right].$$

4. Generate ω_A

$$[\omega_A \mid \{Z_s, \beta_{i,s}, \Lambda_s\}] \propto [\{\beta_{i,s}\} \mid \{Z_s, \Lambda_s\}, \theta][\omega_A],$$

where $[\{\beta_{i,s}\} \mid Z_s, \theta, \Lambda_s]$ is the product of multivariate normal densities, $\prod_{s=1}^{S}\prod_{i=1}^{n}[\beta_{i,s} \mid Z_s, \theta, \Lambda_s]$.

An Independent M-H algorithm is used to generate the draws of ω_τ. Generate draws independently from the Uniform(0,1) distribution. Let $\omega_A^{(k)}$ be the *k*th draw, and $\omega_A^{(k+1)}$ be the next draw. Let

$$\theta^{(k)} = (\mu, \kappa_A, \nu_A, \kappa_A\omega_A^{(k)}, \nu_A\omega_A^{(k)}, \theta_P, \theta_C, \theta_T)'.$$

The probability of accepting the new draw $\omega_\tau^{(k+1)}$ is given by

$$\min\left[\frac{[\{\beta_{i,s}\} \mid \{Z_s, \Lambda_s\}, \theta^{(k+1)}]}{[\{\beta_{i,s}\} \mid \{Z_s, \Lambda_s\}, \theta^{(k)}]}, \quad 1\right].$$

5. Generate $\{\Lambda_s$, *s*=1, …, S$\}$ from the Inverted Wishart Distribution

$$[\Lambda_s \mid \{\beta_{i,s}\}, Z_s, \theta] \propto [\{\beta_{i,s}\} \mid Z_s, \theta, \Lambda_s][\Lambda_s] \sim \text{Inverted Wishart } (\eta_s, \Delta_s)$$

$$\eta_s = \eta_{0,s} + n_s, \quad \Delta_s^{-1} = \Delta_{0,s}^{-1} + \sum_{i=1}^{n_s}(\beta_{i,s} - Z_s\theta)(\beta_{i,s} - Z_s\theta)'.$$

300,000 iterations are run and every 30th- iteration is kept. Posterior means and variances are calculated using draws after the burn-in period of the first 60,000 iterations.

APPENDIX C: Prediction Validation Results

Actual $\bar{y}_3$	$\hat{\bar{y}}_3$ from Design I	$\hat{\bar{y}}_3$ from Design II	$\hat{\bar{y}}_3$ from Design III	$\hat{\bar{y}}_3$ from Design IV
1.287	1.885 (0.224)	2.001 (0.219)	1.653 (0.197)	1.756 (0.196)
0.681	0.281 (0.146)	0.295 (0.135)	0.238 (0.139)	0.071 (0.138)
1.011	0.863 (0.177)	0.743 (0.171)	0.935 (0.156)	0.828 (0.163)
1.383	1.934 (0.193)	1.940 (0.190)	1.668 (0.172)	1.885 (0.181)
4.043	4.326 (0.227)	4.307 (0.240)	4.541 (0.219)	4.437 (0.209)
3.362	3.793 (0.197)	3.799 (0.208)	3.860 (0.190)	3.810 (0.182)
5.000	5.447 (0.220)	5.443 (0.228)	5.290 (0.207)	5.624 (0.197)
3.351	3.843 (0.192)	3.738 (0.212)	3.876 (0.187)	3.939 (0.174)
5.181	4.898 (0.214)	4.940 (0.235)	5.134 (0.204)	5.004 (0.209)
5.745	5.969 (0.214)	6.137 (0.216)	5.868 (0.197)	6.062 (0.196)
5.330	4.948 (0.218)	4.879 (0.236)	5.150 (0.207)	5.133 (0.206)
4.436	4.415 (0.187)	4.371 (0.209)	4.469 (0.184)	4.506 (0.178)

Table C.1: Predictions of mean profile ratings in holdout study version 1

Actual $\bar{y}_3$	$\hat{\bar{y}}_3$ from Design I	$\hat{\bar{y}}_3$ from Design II	$\hat{\bar{y}}_3$ from Design III	$\hat{\bar{y}}_3$ from Design IV
0.979	0.814	0.804	0.919	0.698
	(0.193)	(0.190)	(0.180)	(0.180)
0.511	0.281	0.295	0.238	0.071
	(0.146)	(0.135)	(0.139)	(0.138)
1.043	1.934	1.940	1.668	1.885
	(0.193)	(0.190)	(0.172)	(0.181)
0.543	0.330	0.234	0.254	0.201
	(0.115)	(0.115)	(0.112)	(0.113)
5.532	5.397	5.504	5.275	5.495
	(0.235)	(0.240)	(0.216)	(0.203)
3.755	3.793	3.799	3.860	3.810
	(0.197)	(0.208)	(0.190)	(0.182)
4.426	4.376	4.246	4.557	4.566
	(0.227)	(0.240)	(0.213)	(0.204)
5.383	5.447	5.443	5.290	5.624
	(0.220)	(0.228)	(0.207)	(0.197)
5.277	4.898	4.940	5.134	5.004
	(0.214)	(0.235)	(0.204)	(0.209)
6.468	5.969	6.137	5.868	6.062
	(0.214)	(0.216)	(0.197)	(0.196)
5.277	4.948	4.879	5.150	5.133
	(0.218)	(0.236)	(0.207)	(0.206)
4.415	4.415	4.371	4.469	4.506
	(0.187)	(0.209)	(0.184)	(0.178)

Table C.2: Predictions of mean profile ratings in holdout study version 2

Actual $\bar{y}_3$	$\hat{\bar{y}}_3$ from Design I	$\hat{\bar{y}}_3$ from Design II	$\hat{\bar{y}}_3$ from Design III	$\hat{\bar{y}}_3$ from Design IV
1.354	1.885	2.001	1.653	1.756
	(0.224)	(0.219)	(0.197)	(0.196)
0.594	0.281	0.295	0.238	0.071
	(0.146)	(0.135)	(0.139)	(0.138)
1.177	0.863	0.743	0.935	0.828
	(0.177)	(0.171)	(0.156)	(0.163)
1.208	1.934	1.940	1.668	1.885
	(0.193)	(0.190)	(0.172)	(0.181)
0.583	0.330	0.234	0.254	0.201
	(0.115)	(0.115)	(0.112)	(0.113)
4.250	4.326	4.307	4.541	4.437
	(0.227)	(0.240)	(0.219)	(0.209)
3.490	3.793	3.799	3.860	3.810
	(0.197)	(0.208)	(0.190)	(0.182)
5.646	5.447	5.443	5.290	5.624
	(0.220)	(0.228)	(0.207)	(0.197)
5.135	4.898	4.940	5.134	5.004
	(0.214)	(0.235)	(0.204)	(0.209)
6.583	5.969	6.137	5.868	6.062
	(0.214)	(0.216)	(0.197)	(0.196)
5.427	4.948	4.879	5.150	5.133
	(0.218)	(0.236)	(0.207)	(0.206)
5.063	4.415	4.371	4.469	4.506
	(0.187)	(0.209)	(0.184)	(0.178)

Table C.3: Predictions of mean profile ratings in holdout study version 3

References

Atkinson, A.C. and A.N. Donev (1992), *Optimum Experimental Designs,*, Clarendon Press, Oxford.

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

Chaloner, Kathryn and Isabella Verdinelli (1995), "Bayesian Experimental Design: A Review," *Statistical Science*, 10 (3), 273–304.

Cooke, Alan D.J., Chris Janiszewski, Marcus Cunha Jr., Suzanne A. Nasco, and Els De Wilde (2004), "Stimulus Context and the Formation of Consumer Ideals," *Journal of Consumer Research*, 31 (June), 112–124.

Creyer, Elizabeth and William T. Ross (1988), "The Effects of Range-Frequency manipulations on Conjoint Importance Weight Stability," *Advances in Consumer Research*, 15, 505–509.

Currim, Imran S., Charles B. Weinberg, and Dick.R. Wittink (1981), "Design of Subscription Programs for a Performing Arts Series," *Journal of Consumer Research*, 8 (June), 67–75.

Gamerman, Dani (1997), *Markov Chain Monte Carlo*, Chapman & Hall.

Han, Cong and Kathryn Chaloner (2004), "Bayesian Experimental Design for Nonlinear Mixed-effects Models with Applications to HIV Dynamics," *Biometrics*, 60, 25–33.

Huber, Joel, John W. Payne, and Christopher Puto (1982), "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis," *Journal of Consumer Research*, 9 (June), 90–98.

Janiszewski, Chris and Donald R. Lichtenstein (1999), "A Range Theory Account of Price Perception," *Journal of Consumer Research*, 25 (March), 353–368.

Kahn, Barbara, William L. Moore, and Rashi Glazer (1987), "Experiments in Constrained Choice," *Journal of Consumer Research*, 14 (June), 96–113.

Krumhansl, Carol L. (1978), "Concerning the Applicability of Geometric Models to Similarity Data: The Interrelationship Between Similarity and Spatial Density," *Psychological Review*, 85 (5), 445–463.

Kuhfeld, Warren F (2005), "Experimental Design, Efficiency, Coding and Choice Designs," Tech. rep., SAS TS-722C, http://support.sas.com/techsup/tnote/tnote_stat.html.

Lenk, Peter J., Wayne S. Desarbo, Paul E. Green, and Martin R. Young (1996), "Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs," *Marketing Science*, 15 (2), 173–191.

Liu, Qing, Angela M. Dean, and Greg M. Allenby (2007), "Optimal Experimental Designs for Hyper-parameter Estimation in Hierarchical Linear Models, " http://www.stat.osu.edu/~amd/dissertations.html, *Submitted for publication*.

Lynch, John G., Dipankar Chakravarti, and Anusree Mitra (1991), "Contrast Effects in Consumer Judgments: Changes in Mental Representations or in the Anchoring of Rating Scales?" *Journal of Consumer Research*, 18 (December), 284–297.

Mentre, France, Alain Mallet and Doha Baccar (1997), "Optimal Design in Random-Effects Regression Models," *Biometrika*, 84 (2), 429–442.

Parducci, Allen (1965), "Category Judgment: A range-frequency model," *Psychological Review*, 72, 407–418.

——— (1974), "Contextual Effects: A Range-Frequency Analysis," in *Handbook of Perception*, eds. Edward Carterette and Morton Friedman, New York: Academic Press, 127-141.

——— (1982), "Category Ratings: Still more contextual effects," in *Social Attitudes and Psychophysical Measurement*, eds. Wegener B., Hillsdale, NJ: Erlbaum, 89-105.

Parducci, Allen and Douglas H. Wedell (1986), " The category effect with rating scales: Number of categories, number of stimuli, and method of presentation," *Journal of Experimental Psychology: Human Perception and Performance*, **12**, 496-516.

Rossi, Peter E., Greg M. Allenby, and Robert McCulloch (2005), *Bayesian Statistics and Marketing*, John Wiley and Sons, Ltd.

Sandor, Zsolt and Michel Wedel (2002), "Profile Construction in Experimental Choice Designs for Mixed Logit Models," *Marketing Science*, 21 (4), 455–475.

Simonson, Itamar and Amos Tversky (1992), "Choice in Context: Tradeoff Contrast and Extremeness Aversion," *Journal of Marketing Research*, 29 (3), 281–295.

Steenkamp, Jan-Benedict E.M. and Dick R. Wittink (1994), "The Metric Quality of Full-Profile Judgments and the Number-of-attribute-levels Effect in Conjoint Analysis," *International Journal of Research in Marketing*, 11, 275–286.

Tod, Michel, France Mentre, Yann Merle, and Alain Mallet (1998), "Robust Optimal Design for the Estimation of Hyper-parameters in Population Pharmacokinetics," *Journal of Pharmacokinetics and Biopharmaceuics*, 26, 689–716.

Tversky, Amos (1977), "Features of Similarity," *Psychological Review*, 84 (3), 327–352.

Verlegh, Peter W.J., Hendrik N.J. Schifferstein, and Dick R. Wittink (2002), "Range and Number-of-Levels Effects in Derived and Stated Measures of Attribute Importance," *Marketing Letters*, 13 (1), 41–52.

Wedell, Douglas.H., Allen Parducci and Diana Roman (1989), "Student perceptions of fair grading: A range-frequency analysis," *American Journal of Psychology*, **102**, 233- 248.

Wittink, Dick R., Joel Huber, Peter Zandan, and Richard M. Johnson (1992), "The Number of Levels Effect in Conjoint: Where Does It Come From, And Can It Be Eliminated?" in *Sawtooth Software Conference Proceedings*.

Wittink, Dick R., Lakshman Krishnamurthi, and Julia B. Nutter (1982), "Comparing Derived Importance Weights Across Attributes," *Journal of Consumer Research*, 8, 471–474.

Wittink, Dick R., Lakshman Krishnamurthi, and David J. Reibstein (1990), "The Effect of Differences in the Number of Attribute Levels on Conjoint Analysis," *Marketing Letters*, 1 (2), 113–123.

Wittink, Dick R., William G. McLauchlan, and P.B. Seetharaman (1997), "Solving the Number-of-Attribute-Levels Problem in Conjoint Analysis," in *Sawtooth Software Conference proceedings*.