MICHIGAN
ROSS SCHOOL OF BUSINESS

Working Paper

Split Questionnaire Design

Feray Adigüzel
University of Groningen, The Netherlands

Michel Wedel
Stephen M. Ross School of Business
at the University of Michigan

Ross School of Business Working Paper Series
Working Paper No. 966
January 2005

UNIVERSITY OF MICHIGAN

Split Questionnaire Design

Feray Adıgüzel

University of Groningen[1]

P.O. Box 800 9700AV Groningen, The Netherlands

And

Michel Wedel

Stephen M. Ross School of Business

At The University of Michigan

701 Tappan Street, MI 48109 Ann Arbor, USA.

[1] The authors gratefully acknowledge the support of the Netherlands National Science Foundation, Grant number 490-24-121, and the support of the University of Michigan Business School, the first author was a visiting scholar there during part of the work on this project (e-mail:adiguzel@bus.umich.edu).

Split Questionnaire Design

Abstract

Instead of the heuristic randomization methods to design split questionnaires that are currently used in applied and academic research, we develop a methodology to design the split questionnaire to minimize information loss. Because the number of possible questionnaire designs is exponential in the number of questions, we apply the Modified Federov algorithm, using Kullback Leibler Distance as a design criterion, to find the optimal splits. First of all, we illustrate the efficiency of the Modified Federov Algorithm on a small synthetic questionnaire, which enables the enumeration of all possible designs for comparison. Second, we compare the efficiency of split questionnaires generated with the proposed method to multiple matrix sampling (randomly generated designs) and a heuristic procedure based on principal components analysis, on synthetic and empirical data. We generate split questionnaire designs selecting either entire blocks of questions (between-block design) or sets of questions in each block (within-block design). Finally, we illustrate that due to reduced respondent burden the quality of data using split designs increases, compared to a full questionnaire in a field study.

INTRODUCTION

Market researchers have traditionally collected consumer information on preferences, attitudes, consumption contexts and lifestyles, by means of often very long questionnaires. In doing so, they need to make tradeoffs between reasonable survey length and the value and quality of additional information. Questionnaire length is a concern since it affects the quality of the data collected in several ways (Berdie 1989). Long questionnaires lead to higher non-response, item non-response and early break-off rates. They also cause an increase in the use of undesired response styles, increased time to collect the data, and respondent fatigue and boredom. Survey respondents are reported to become fatigued and irritable when questioned last for more than twenty minutes. Many studies indicate that longer questionnaires have lower response rates than shorter ones (Adams and Gale 1982; Bean and Roszkowski 1995; Dillman 1991; Dillman, Sinclair, and Clark 1993; Heberline and Baumgartner 1978; Roszkowski and Bean 1990).

Motivation

We propose a method to design split-questionnaire surveys as an effective tool to reduce respondent burden without sacrificing the inferential content of the data. Although Good (1969, 1970) already called for the development of split-questionnaire methods to collect survey data more efficiently, in the next thirty-five years no systematic research on how to best design split-questionnaires seems to have been done. Two decades ago, Herzog and Bachman (1981) advised that a researcher who needs to use a long questionnaire might be well advised to split the material into at least two parts and administer those parts in different orders to different random subsets of the sample. In their split questionnaire survey design the original questionnaire is divided into sub-components and subjects respond to a randomly selected set of components only. A similar idea of designing randomly split questionnaires is applied in what has been called "Time sampling". Here, questions are administered in a randomly rotated fashion to different parts of the panel in different episodes (Sikkel and Hoogendoorn 1995). Incomplete designs in educational testing are based on a similar approach. In test construction the researcher administers subsets of the total available item pool to the available subjects. The matrix sampling design is used for that purpose (Shoemaker 1973, Thayer 1983), in which a test instrument is divided in sections, and groups of sections are administered to subjects in a randomized fashion.

Each of those previous studies has thus used a randomization approach to design split questionnaires. The important question remains how to optimally split the questionnaire, such that the least information is lost. Currently, no methods have been published to address that problem, and here lies the contribution of the present study. Raghunathan and Grizzle (1995) mention that ad-hoc splitting strategies may depend on the purpose and the contents of the survey, contextual placement of certain items, and the partial correlation coefficients of the items. These correlations may be readily available in tracking or syndicated studies, because here the researcher knows which (groups of) variables are correlated, from their previous measurements. In cross-sectional studies prior knowledge about inter-relationships between variables can be obtained from a pilot study. However, even when such prior information is available, the construction of a split-questionnaire design such that a minimum amount of information is lost is a challenging task. Since the number of possible split-questionnaire designs is exponential in the number of questions, it is not feasible to consider all possible splits in designing a questionnaire for real-life applications. Therefore, we suggest, in line with previous practice in marketing research, to utilize the natural structure of the questionnaire, in which questions are placed in blocks. Mostly, several questions measuring for example one particular attitudinal or lifestyle trait are administered as a group or block. We use this block-structure to generate split-questionnaire designs in two different ways: selecting entire blocks of questions, which we call a “Between-block design”, or selecting questions in each block, which we call a “Within-block design”. In the between-block design, a “split” comprises of the allocation of selected blocks of questions and respondents answer all questions in these blocks; in the within-block design, a split comprises of sets of selected questions in each of the blocks and respondents answer only those questions in each block[2]. For the first method, given the coherent interpretation of the questions in one block, the problem then simplifies to how these blocks should be administered to respondents in an optimal way. On the other hand, for the within-block design we need to choose questions in each block optimally. The choice between the within-block and the between-block design should be based on substantive issues, as well as statistical properties of the two types of designs, as will become clear in the sequel of this paper. We focus on the problem of how to best develop a split-questionnaire and propose a method to optimally choose the splits (a set of blocks of questions or questions in each block offered to a respondent).

Outline of the paper

The main contribution of this paper is to propose a method to design split-questionnaires. We apply the Modified Federov Algorithm to find the optimal design from all possible designs because of its fast and reliable properties. This method has been previously applied in a different context in the design of conjoint experiments (Kuhfeld, Tobias and Garratt 1994). We propose to use Kullback-Leibler (KL) Distance between the complete and split-questionnaire data as an optimization criterion. The algorithm searches the candidate splits for the split that is optimal in terms of the criterion. As explained above, we study both between-block and within-block split

2 We acknowledge one anonymous reviewer for suggesting the within-block design to us.

questionnaire designs. The split questionnaire, once administered, results in data missing by design, which may result in lack of identification of all parameters from the observed data (Little and Rubin 1997; Rassler 2002). Specific overlap of the splits of the questionnaire may help to avoid that identification problem. We explain how to construct identified split-questionnaire designs, and how to impute the missing data with the Gibbs sampler. Using a small simulated questionnaire we enumerate all possible designs and compare that with the result of our design generating algorithm, which reveals that it recovers the optimal split in all cases. We compare the efficiency of split questionnaires generated with our procedure to (random) matrix sampling designs on synthetic data. In practice, market research companies design split-questionnaires by randomly choosing blocks, or questions in each block. These methods are similar to the multiple matrix sampling techniques used in testing theory (Shoemaker 1973), and will therefore constitute an appropriate benchmark.

We then apply our approach to data obtained from a questionnaire on web attitudes and perceptions (Novak, Hoffman, and Yung 2000) to assess the performance of optimal between- and within-block designs empirically, and compare them to matrix sampling designs and heuristic designs constructed based on a principal components analysis of pilot data. We investigate the sensitivity of the optimal split questionnaire designs to changes in the prior parameters from the pilot study. Finally, we investigate the extent to which the proposed split questionnaire design method may result in better data quality than the complete questionnaire, by studying respondent burden, boredom, and fatigue in a field application of the web-attitude questionnaire. Our conclusion is that optimally splitting questionnaires is worth consideration for improving the efficiency of questionnaires and the resulting data quality.

The subsequent sections are organized as follows: Section 2 examines issues in designing a split-questionnaire. In section 3, the design criterion is introduced; the Modified Federov Algorithm and the construction of identified split- designs are explained. In section 4, we discuss multiple imputations of the missing data and the estimation of the fraction of missing information. Section 5 provides a simulation study to investigate the performance of the proposed split questionnaire design method, section 6 provides the empirical application and section 7 summarizes the field study. Finally, in section 8 the results of this research are discussed and concluding remarks are offered.

CONSTRUCTING THE SPLIT QUESTIONNAIRES

Finding an optimal design for a split-questionnaire involves finding the configuration of question sets (i.e. those questions given to one respondent, or a "split") such that a minimum amount of information is lost as compared to the complete questionnaire. The design of a split questionnaire, as we propose it, involves two steps. First, one needs to assign questions to blocks with homogeneous content. Second, one needs to allocate either selected blocks to splits, or selected questions within blocks to splits, resulting in between- and within-block designs, respectively. In the first step one wants to keep thematically closely related questions in the same block[3]. Raghunathan and Grizzle (1995) call this the contextual placement of questions. We start from the assumption that the questionnaire already consists of a number of blocks with questions that need to be kept together, and we will utilize that natural structure of the questionnaire**.** Our approach is thus very suitable for questionnaires comprising of items to measure several multi-

[3] A block structure, if not available a-priori, can be generated using cluster analysis of a pilot with the full questionnaire (Rassler 2002).

item constructs. These are very common in marketing research. Each split questionnaire design is defined by three sets of parameters: the number of splits, the number of blocks/questions per split, and the sampling fraction responding to each split. In this study we investigate the first two parameters and assume throughout that splits are distributed evenly and at random to respondents. We propose to choose splits from all possible combinations of blocks (between-block designs) or from all possible combinations of questions in each block (within-block designs), using the Kullback-Leibler distance as a measure of information loss, computed from prior parameter estimates.

MEASURING INFORMATION LOSS

Optimal Split Questionnaires Using KLD

We use the Kullback-Leibler (KL) measure, the distance between two probability models, to choose the best among all possible designs. The KL-distance was developed by Kullback and Leibler (1951) from "information theory''. Here, it is first applied to design construction. The KL distance defines the distance between the probabilistic models *f* and *g* for as the (usually multi-dimensional) integral:

(1) $$I(f,g) = \int f(y)\log\left(\frac{f(y)}{g(y\,|\,\theta)}\right)dy$$

I(f,g) is the "information'' lost when *g* is used to approximate *f*. An equivalent interpretation of minimizing I(f,g) is finding an approximating model that is the shortest distance away from "the truth". If f(y) and g(y|θ) are multivariate normal distributions with common variance-covariance matrix then the Kullback-Leibler distance reduces to the Mahalanobis distance (Bar-Hen and Daudin 1995), which is frequently used as a distance measure in the literature.

We assume that the optimization of the split-questionnaire design (SQD) is done under one external constraint fixed by the researcher, which is the total number of splits (*K*) desired. We assume that the researcher knows this number from prior considerations, or that issues related to the implementation of the questionnaire dictate it. The optimization can also accommodate any other practical constraint, such as one that induces respondents to answer a fixed number of (blocks of) questions, i.e. each candidate split should contain a predetermined number of blocks. These constraints are illustrated below. After generating *K* splits and evenly distributing these splits to respondents, the Kullback-Leibler distance is calculated. In our notation, *K* denotes the total number of splits, *N* is the number of respondents, *B* is the number of blocks, Q_b is the number of questions in block b, *Q* is the total number of questions, $(\sum_{b=1}^{B} Q_b = Q)$, *Y* is the data-matrix containing the answers of the respondents and *D* is the questionnaire design matrix with 0/1 entries (i.e. a fully observed matrix of indicators whose elements are zero or one depending on whether the corresponding elements of Y are missing or observed):

$$d_{ij} = \begin{cases} 1 & \text{if question j is given to respondent i} \\ 0 & \textit{otherwise} \end{cases}$$

Now f(Y|D) is the likelihood of the incomplete data with respect to the split questionnaire design matrix and f(Y) is likelihood of the data with respect to the complete questionnaire. The Kullback-Leibler distance between the complete data likelihood f(Y) and the split data likelihood f(Y|D) is defined as:

(2) $$\mathrm{KL(D)} = \int \mathrm{f(Y)} \ln\left[\frac{\mathrm{f(Y)}}{\mathrm{f(Y \mid D)}}\right] \mathrm{dY},$$

$$= \mathrm{E} \ln[\mathrm{f(Y)}] - \mathrm{E} \ln[\mathrm{f(Y \mid D)}],$$

where each expectation is with respect to the true distribution f(Y), where $Y_{N\times Q} = [Y_1, Y_2, \ldots, Y_Q]$. The most efficient questionnaire design (D) minimizes KL(D). The first term on the right hand side in the equation for KL(D) is the same for each possible design since it is derived from the complete questionnaire. Consequently, maximization of the second term on the right hand side suffices. Since f(Y) is the same for each possible design, lnf(Y|D) will be maximized in the sequel. Minimizing the KL distance can be seen as finding the split questionnaire that yields incomplete data that are closest in expectation to the data that would have been obtained with the complete questionnaire.

We will assume the form of lnf(Y|D) to be a multivariate normal, as a function of the parameters μ and Σ, as shown below. In Appendix I we provide an extension of the KL distance for mixed data consisting of continuous and discrete variables using a general location model. But, multivariate normality is often assumed for responses of scales in many marketing surveys, including those measuring attitudes, satisfaction, lifestyles etc. (Huber et al. 1993). In addition, the normal distribution has minimal KL distance to any unknown distribution function (O' Hagan 1994), and in this case minimizing the KL-distance is equivalent to minimizing the Mahalanobis distance.

We have Q-variate normal data $N_Q(\mu, \Sigma)$ with $\mu = (\mu_1, \ldots.., \mu_Q)$ and $\Sigma_{Q\times Q}$. For now, $\mu_{Q\times 1}$ and $\Sigma_{Q\times Q}$ are assumed known. These are considered prior information that can be obtained from past data or through a pilot experiment. The aim is to construct the design using $\mu_{Q\times 1}$ and $\Sigma_{Q\times Q}$ as prior information. Thus, we have the following optimal design criterion:

(3) $$\ln L(Y \mid D, \mu, \Sigma) = \prod_{i=1}^{n} -\left(\frac{p_D}{2}\right)\ln(2\pi) - \frac{1}{2}\ln\left|\Sigma(D)\right| - \frac{1}{2}\left[(Y_{obs} - \mu(D))'\Sigma(D)^{-1}(Y_{obs} - \mu(D))\right]$$

with p_D is the number of parameters under design D, *n* the total number of respondents, $Y_{obs}=Y_{ij}d_{ij}$ the data observed under the split-questionnaire D, and $\mu(D)$ and $\Sigma(D)$ denote the subvector of the mean vector μ and the square submatrix of the covariance matrix Σ which are obtained from complete data estimates from a pilot study, respectively, that pertain to the variables that are observed in design D.

Identification Issues in Constructing SQD

When we construct a split questionnaire design, we should be able to estimate all parameters from the observed incomplete data. We call a design that enables the estimation of all parameters (of the multivariate normal distribution) a fully identified design. Clearly, not all designs are fully identified. We illustrate the identification problem briefly through the following example. Assume we want to estimate the parameters of a multivariate Normal distribution for three blocks, X, Y and Z in a between-block design. However, we have a split A- with only X and Y and a split B- with only X and Z observed together. We have $V(Y,Z) = V(Y,Z \mid X) + V'(X,Y)V(X)^{-1}V(X,Z)$ with V(Y,Z) the covariance matrix of Y and Z, V(X) the covariance matrix of X, and V(Y,Z|X) the covariance matrix of Y and Z conditional on X. We can estimate V(X,Y) from split A, V(X,Z) from split B, and V(X) from both splits, but we cannot only directly estimate V(Y,Z|X) from the available incomplete data. However, if we assume conditional independence of the Y and Z variables given X, we have $V(Y,Z) = V'(X,Y)V(X)^{-1}V(X,Z)$ and can estimate V(Y,Z), since all terms on the right hand side are estimable (see Gilula, McCulloch and Rossi 2004; Rassler 2002; Rodgers 1984). However, if we use this conditional independence assumption in a model for imputing the missing data, this implies that for all parameter estimates or statistics subsequently computed from the imputed data this conditional independence assumption should also hold. That assumption is a strong one, which may limit the usefulness of such split- questionnaire designs in practice.

Rassler (2002) and Gilula, McCulloch and Rossi (2004) suggest (in the context of data-fusion) to use informative priors in the imputation to overcome the identification problem. The use of priors adds information that enables estimation of the parameters that are not identified by the split questionnaire design. The fact that V(Y,Z|X) is inestimable results in non-positive definite variance-covariance matrix V(X,Y,Z), which we can avoid using prior information. If one uses the Gibbs sampler for imputation, as we will below, such prior information also overcomes lack of convergence. Using informative priors for the means and covariance matrix of the normal distribution results in an imputed dataset devoid of conditional independence properties induced by the design, which is highly desirable. Since the design itself is constructed based on such prior information, it is natural to also include that same prior information in imputing the missing data. However, it is even more desirable to address the identification problem by constructing designs that do not suffer from it, which we do below.

If all possible pairs of questions occur in an optimal split questionnaire design, this ensures that all parameters of a multivariate normal distribution are identified and estimable from the observed data. Let us consider the between-block design: if we have a questionnaire with n_B blocks and we impose the constraint of n_S blocks per split, then the number of splits K, for a fully identified design needs to satisfy $\binom{n_B}{n_S} \leq K \leq \frac{n_B(n_B - 1)}{n_S(n_S - 1)}$, where $\binom{n_B}{n_S}$ is the size of the candidate split-set. Note that is a necessary, but not sufficient condition. In practice one can easily check the identification of any design by looking at the (D'D) matrix: only designs with all off-diagonal elements greater than 0 are fully identified designs. In generating constrained split questionnaire designs, we recommend that one only considers fully identified designs by imposing the identification constraint $(d_i' d_j) \neq 0, \quad \forall_{i \neq j}$, and employ the prior information used to construct the design also in imputing the missing values. This is what we will do throughout the remainder of this paper and we recommend it in general as a procedure for constructing split-questionnaires.

Efficient Design Generating Algorithms

In order to find the most efficient K splits out of all possible candidate splits ($N_S = 2^Q$, with Q the number of questions), all $N_D = \binom{2^Q}{K}$ possible designs should be generated and evaluated. In most practical situations, it is not computationally feasible to find the global maximum of ln[f(Y|D)] among all possible K subsets out of 2^Q points, since it is usually not possible to list all N_D-designs because run time is exponential in the number of candidates. As a result, we need to use an efficient design-generating algorithm. Such an algorithm searches among all possible candidate splits for one that improves a given criterion.

We apply the Modified Federov algorithm to obtain the optimal questionnaire design. Kuhfeld, Tobias and Garratt (1994) suggest the Modified Federov algorithm, since it is robust and fast, and apply it in the different context of constructing conjoint designs. For split-questionnaire designs, we begin by building a candidate split-set (C, a $N_S \times Q$ matrix), which is a list of potential splits. If there are N individuals, then N/K individuals will be assigned randomly to each of the K splits. Each alternative split-questionnaire design consists of an N x Q matrix D with K different split patterns. Each entry in the matrix is a 0 or 1, indicating whether a question is included or excluded in that particular split. The starting design is chosen at random. The algorithm exchanges its splits (i.e. each row of the matrix) with the candidates. It finds the best exchange (if one exists) for the first split in the starting design, by sequentially processing the candidates. Then it moves on to the second split, and so on. The first iteration is completed once the algorithm has found the best exchanges for all of the splits (rows) in the starting design. Then, the algorithm moves back to the first split and continues to replace it with each candidate and continues in that fashion until no improvement is possible. To avoid local optima, the whole process is restarted with different (random) starting designs and the best design is selected.

Generating Between-Block and Within-Block Designs

In constructing between-block designs, first of all we generate all 2^B possible splits of blocks. We assume that all questions in one block are assigned to the same respondent. That is, if we have five blocks with four questions and one particular split is 11010, we will use d_{ij}=[1111 1111 0000 1111 0000]. Using the Modified Federov algorithm, we choose K different splits from all candidate splits and obtain the optimal design, ensuring that it is identified by checking off-diagonals of the (D′D) matrix.

Whereas the construction of between-block designs is feasible in this manner that of the within-block design is not, in most practical situations. Because of the enormous size of the design space for within-block designs, even generating the candidate split set is nontrivial. Therefore, we choose questions from each block using a “greedy” approach, as follows. Instead of optimizing of the full within-block split design, we generate splits for each block sequentially. For block B there are 2^{Q_B} possible splits with Q_B the number of questions. We first find the optimal K splits in the first block using the Modified Federov algorithm as described above, assuming the other blocks are complete. Then, we find the optimal splits in the second block searching across the candidate splits, given the optimal splits of the first block and assuming the remaining blocks are complete. We continue this procedure by sequentially passing through the remaining blocks, finding the optimal splits for each block given the optimal designs of the previous blocks and assuming the remaining blocks complete. Thus, we obtain an optimal, or at least an improved, solution by sequentially producing locally optimal designs within each block. Unfortunately, it proves not easy to produce fully identified within-block designs using the "greedy" approach just described. We therefore produce locally identified designs by checking the $D_b'D_b$ matrix of each block b separately. However, this does not guarantee the appearance of all question-pairs in the complete design, which is needed for the design to be fully identified. Thus, the constructed within-block split questionnaire designs are neither fully identified nor globally optimal, but, are still more efficient than designs constructed by choosing questions within each block at random or with heuristic procedures. We investigate this in detail below.

MULTIPLE IMPUTATIONS WITH GIBBS SAMPLING

The within- and between-block split questionnaire designs produce datasets with intentionally missing data. To obtain complete data, instead of using a single imputation, which ignores uncertainty due to imputation and therefore underestimates the variability of the resulting estimates (Rubin 1987), we use Bayesian proper multiple imputations by drawing values of missing data (Y_{mis}), and μ and Σ from their full conditional posterior distributions using Gibbs sampling (Gelfand and Smith 1990). We use informative priors, μ_{pr} and Σ_{pr}, obtained from the full questionnaire in a pilot study, with n_0 and ρ the prior number of observations and degrees of freedom on which the μ_{pr} and Σ_{pr} are based, respectively. Let $\Sigma_{obs,obs}$, $\Sigma_{mis,mis}$, and $\Sigma_{mis,obs}$ denote the sub-matrices of Σ formed by the indices corresponding to the observed and missing Y values; μ_{obs}, μ_{mis} denote the corresponding sub-vectors of μ. The conditional distribution of Y_{mis}, given

Y_{obs}, μ_m, and Σ is normal with mean $\mu_{mis} + \Sigma_{obs,mis}\Sigma^{-1}_{obs,obs}(Y_{obs} - \mu_{obs})$ and variance $\Sigma_{mis,mis} - \Sigma_{obs,mis}\Sigma^{-1}_{obs,obs}\Sigma_{mis,obs}$. The Gibbs sampler iterates between:

1) draw $Y^{(t+1)}_{mis}$ given μ_0, Σ_0, and Y_{obs}:

(4) $$Y^{(t+1)}_{mis} \mid Y_{obs} \sim MVN\left(\mu_{mis} + \Sigma_{obs,mis}\Sigma^{-1}_{obs,obs}(Y_{obs} - \mu_{obs}); \Sigma_{mis,mis} - \Sigma_{obs,mis}\Sigma^{-1}_{obs,obs}\Sigma_{mis,obs}\right),$$

2) draw $\Sigma^{(t+1)}$ given $\mu^{(t)}$ and $Y^{(t+1)} = (Y_{obs}, Y^{(t+1)}_{mis})$ from:

(5) $$\Sigma^{(t+1)} \mid Y \sim IW(n_{obs} + \rho, (n_{obs} - 1)S + \rho \times \Sigma_{pr} + S_m),$$

where S is the sample covariance matrix and $S_m = \frac{n_{obs} \times n_0}{(n_{obs} + n_0)}(\bar{y} - \mu_{pr})(\bar{y} - \mu_{pr})'$,

3) draw $\mu^{(t+1)}$ given $\Sigma^{(t+1)}$ and $Y^{(t+1)}=(Y_{obs}, Y^{(t+1)}_{mis})$ from

(6) $$\mu^{(t+1)} \mid \left(\Sigma^{(t+1)}, Y\right) \sim N\left(\frac{1}{n_{obs} + n_0}(n_{obs}\bar{y} + n_0\mu_{pr}), \frac{1}{n_{obs} + n_0}\Sigma^{(t+1)}\right).$$

The Gibbs sampler is easy to implement and enables quick imputation of the missing values.

Estimation of the Fraction of Missing Information

The incomplete data generated through the split-questionnaire design contain less information on the parameters than the complete data. We estimate the fraction of missing information of the parameters using the missing information principle (Orchard and Woodbury 1972). Since the complete data information is the sum of the observed data information and the missing data information, we can write:

(7) $$\frac{1}{V(\hat{\theta})} = \frac{1}{V(\hat{\theta}_{obs})} + \left(\frac{1}{V(\hat{\theta})} - \frac{1}{V(\hat{\theta}_{obs})}\right).$$

Here $V(\hat{\theta})$ is the complete information on θ estimated from the Fisher information matrix. $V(\hat{\theta}_{obs})$ is the expected observed data information, which we estimate after the multiple imputation of the missing data with the Gibbs sampler. If we divide both sides by the missing information and take the fraction of missing information (γ) to be equal to the missing information divided by the complete information, we obtain:

$$\gamma = \frac{\left(\frac{1}{V(\hat{\theta})} - \frac{1}{V(\hat{\theta}_{obs})}\right)}{\frac{1}{V(\hat{\theta})}} \tag{8}$$

This quantity shows how much information there is in the data on the parameters in question, and can be used as a statistic to evaluate the efficiency of split-questionnaire designs.

SIMULATION STUDIES

Before we extensively investigate the performance of split questionnaire designs on empirical data below, we first illustrate them on simulated data. We conduct two simulation studies, focusing on between-block designs. First, we investigate the performance of the Modified Fedorov algorithm in identifying the optimal design. Second, we compare optimal split questionnaire designs to matrix sampling designs.

We construct a split questionnaire design that is small enough to enumerate all possible designs, which makes it possible to investigate the performance of the Modified Federov Algorithm in finding the optimal design. Let Y_{ij} denote the answer of respondent $i \in \{1, \ldots., N\}$ to question $j \in \{1, \ldots\ldots., Q\}$, which forms the complete data matrix Y. We assume a between-block design, with B = 5 blocks and each block containing Q_b = 4 questions, so that in total we have twenty questions. We generate Y from a multivariate normal distribution with given $\mu_{Q\times1}$ and $\Sigma_{Q\times Q}$. The matrix X is an $N_S \times B$ matrix containing N_S possible or candidate splits, 1 denoting an included block and 0 denoting an excluded block. There are 32 candidate split points contained in the matrix *X*, but unrealistic or undesirable combinations such as one where none of the questions is asked (a row with only zeros in the design matrix *X*) or where just one block of questions is asked, are excluded, as indicated in the candidate split set shown in Table 1. Even under the external constraint that fixes the number of desired splits (*K*), there are many possible designs. For example, there are in total 5311735 (= 26!/(16!10!)) different designs for K = 10 splits. We choose *K* splits from the candidate split matrix in Table 1, and distribute these splits evenly to one hundred subjects. We do this both with the Modified Fedorov algorithm and through complete enumeration. The matrix *D* contains the design with the *K* splits. We eliminate the responses of the subjects from the complete data matrix (Y) according to the split design (D) and compute the KL distance. We choose the SQD design with the maximum lnf(Y|D) among all possible designs as the optimal design. We investigate three different numbers of desired splits: K = 5, K = 10 and K = 15.

[INSERT TABLE 1 ABOUT HERE]

The time that the Modified Federov Algorithm needed to find the optimal questionnaire design with K=5, 10 or 15 splits is compared to that for complete enumeration in Table 2. All calculations are done with a Pentium 3 computer, using the GAUSS software. For the Federov algorithm, we used 10 iterations, and 1000 different random starts. All 1000 random starts produced the same optimal design in all three cases in 1/10th or less of the computation time of complete enumeration, as shown in Table 2. This indicates that the performance of the Federov Algorithm as applied to the problem of split questionnaire design is highly satisfactory.

[INSERT TABLE 2 ABOUT HERE]

We now illustrate the performance of optimal between-block split questionnaire designs (SQD) relative to matrix sampling designs (MSD) in a second simulation study (within-block designs are investigated more extensively in the empirical application below). We have six blocks and five questions per block. We optimally design the questionnaire and impute the resulting missing data with the Gibbs sampler. We investigate constrained and unconstrained between-block designs, with 5 or 10 splits. To assess the performance of the proposed method, next to the fraction of missing information, we compute the KL-distance and the Bayes information criterion (BIC), where BIC=$-2\times\ln f(Y|D)+\ln(N)\times 2$. Further, we calculate the mean absolute deviation (MAD) and the root mean square error (RMSE) of the estimates of variance and covariance parameters for the SQD and the MSD relative to the complete data (the optimal design procedure improves efficiency and thus affects only variance and covariance estimates). The results are shown in Table 3. We obtain better values for the BIC- and KL- statistics and less missing information for the SQD as compared to the MSD. Parameter estimates are also closer to the true values for the SQD: the MAD is equal to 3.143 for 10 splits and 2.817 for 5 splits while these values are equal to 3.730 and 3.210 for the matrix sampling design. The missing information for the unconstrained split designs is 24% (ten splits) and 27% (five splits), and 22% and 29%, for constrained split designs, respectively, when we eliminate 50-60% of the questions. In contrast, the fraction of missing information for the MSD is consistently higher. Since these results support the performance of the SQD, we investigate its performance in an empirical setting in the next section.

[INSERT TABLE 3 ABOUT HERE]

EMPIRICAL DATA APPLICATION

We apply our procedure to a previously published empirical dataset obtained with the "Project 2000 Ninth GVU Survey Web Attitude and Perceptions Questionnaire[4]", which assesses on how people use the Web and their attitudes towards using it (Novak, Hoffman, and Yung, 2000). This type of survey, applied repeatedly to the same panel for purposes of tracking consumer attitudes and behavior, may benefit from the application of split questionnaire designs since it is conducted on a regular basis with an almost identical structure. Although this particular application is less than ideal to illustrate the performance of SQD, since the questionnaire is relatively short, we consider the use of a published questionnaire and publicly available data attractive. There are sixty-five questions, grouped into nine blocks according to content. The first block contains five questions about the role of the Web in life, the second block consists of eight questions on feelings while using Web, the third block is composed of five questions related to the Web activities, there are seven questions in the fourth block about perceptions on using the Web, the fifth block consists of seven questions about attitudes on using the Web, the sixth block contains eight questions about people feelings towards using the Web, the seventh and eighth block comprise of respectively ten and nine questions about attitudes and perceptions and the last block contains questions on flow and usage of Web information. The questions are assessed on 9-point Likert scales and are considered to be continuous and normally distributed for the purposes of the present study.

Data are available for two waves of the study conducted in two consecutive years. We use these as initialization and validation data, containing 500 and 1150 respondents, respectively. All data are complete. The advantage of having access to complete data is that it allows us to

[4] http://elab.vanderbilt.edu/research/topics/flow/project2000.gvu9.htm

assess the performance of the SQD. A disadvantage of using such complete data is that we may underestimate the effect of the split questionnaire design, since we do not procure the advantages of improved quality of the responses due to reduced respondent burden. Therefore we also construct a field study with this questionnaire that we report on in the final part of this paper. The initialization data are derived from a first wave of the survey, which we use for creating the split questionnaire. From the initialization data we calculate the complete data parameter estimates. This enables us to obtain the design, using the Federov algorithm to minimize the Kullback-Leibler distance. We investigate the following designs, where all designs in this study are constructed to be fully identified:

a) Optimal split questionnaire (SQD) and matrix sampling designs (MSD),
b) Designs with five or ten splits,
c) Between-block and within-block designs,
d) Unconstrained or constrained designs.

We consider the MSD (matrix sampling design) as a benchmark for the between-block design. For the within-block SQD, we use as benchmarks a random questionnaire design (RQD, in which questions within blocks are randomly assigned) as well as an ad-hoc procedure based on a principal components analysis of the items, as explained in more detail below. We use about the same total number of questions in all designs. We generate the MSD by randomly choosing five or ten splits from the candidate split matrix and evenly distributing them among respondents, eliminating responses from the complete data matrix Y according to the design in question. For the RQD we apply the same procedure for each block separately, each time randomly selecting splits from the candidate split set. Since we have access to the complete data, we apply the constructed designs to those data to generate the missing data pattern. To compare the designs, we compute the KL distance and BIC statistics, the fraction of missing information, and MAD and RMSE, after imputing the missing data with the Gibbs sampler. We use informative priors obtained from the initialization data, for all designs. We run the Gibbs for 3000 iterations and save the last 600 draws from the predictive distribution for Y_{mis} as imputations; iteration plots show that the chains converge well before the end of the burn-in period.

Between-Block Designs

The MAD and RMSE measures shown in Table 4 reveal that the estimated parameters for the optimal SQD design are close to the complete data parameters. For both the five- and ten-split cases, the SQD improves significantly over the MSD, the MAD being 35% and 45% smaller respectively, and RMSE 34% and 45%. The improvement of the optimal designs over the currently used matrix sampling designs is substantial. The reason for the better performance of the five-split design, which results in 32% lower MAD and 31% lower RMSE than the ten-split design, is that the lower number of splits is associated with a smaller percentage of missing questions. For this particular application, the five-split optimal SQD results in a reduction of around 66% of the questions, with only a 14% information loss. With ten splits we obtain a greater reduction in the number of questions as compared to five splits. Here, while the SQD results in a 14% loss of information, for the MSD the fraction of missing information is larger, 18%. The split questionnaires with five and ten splits are provided in Figure 1.

[INSERT TABLE 4 AND FIGURE 1 ABOUT HERE]

In addition, we investigate the case where constraints are imposed on the SQD. In particular, we construct designs in which each split consists of exactly five blocks. We choose this number, since we need at least five splits to generate fully identified designs under the constraint of five blocks per split. We repeat the design construction and imputation procedure on the empirical data, using five and ten splits, fixing each split to contain five blocks. The results are given in Table 4. We focus first on the five-split design. In this case we reduce the number of questions with about 44%, while it was 66% for the unconstrained SQD. As a results, the constrained SQD yields 9% of missing information, while the unconstrained SQD yields 14% of missing information (these numbers are 7% and 14% respectively for the ten-split SQD). The fraction of missing information is also less for the constrained SQD than for the constrained MSD, as expected, but the logL(D) and BIC for the constrained designs are worse than for the unconstrained designs. The RMSE and MAD measures reveal that the SQD estimates are close to those of the complete data, these measures are even smaller that for the unconstrained design. They are better than for the comparable MSD's, although the differences are smaller than for the unconstrained designs. The reason is that the constraints strongly limit the degrees of freedom for improvement over the MSD, since they reduce the size of the candidate split set. The optimal constrained five and ten-split designs are shown in Figure 2.

[INSERT FIGURE 2 ABOUT HERE]

Within-block Designs

Using the prior estimates from the initialization data, we also construct optimal within-block designs by selecting questions within blocks, as described above. We compare the optimal SQD with designs in which the questions within blocks are selected randomly (RQD). To also compare to a stronger benchmark, we construct split designs using principal component analysis (PCA)[5]. We extract five and ten Varimax rotated components to construct the splits. Questions in a block are discarded for a split if they contribute the least variance for that component. Every

[5] We acknowledge an anonymous reviewer for this suggestion.

question was included at least once, and the design has the same number of questions as the SQD and RQD designs.

The results are shown in Table 4. We reduce 41% and 52% of the questions with the five- and ten- split within-block designs. The BIC and KL-distance of the optimal within-block designs are lower than the random design and the principal components design. The optimal within-block designs are also somewhat better in terms of RMSE and MAD of the parameter values, but the differences are not as large as for the between-block designs. The PCA designs are in between the RQD and optimal SQD on these measures. The average percentage of missing information is around 7.8% and 5.6% respectively for the optimal five- and ten-split designs. These numbers are better than for the corresponding random designs, with 8.7% and 6.0% respectively, and for the PCA designs, with 8.4% and 5.8%, respectively. The fraction of missing information for within-block designs, however, is substantially lower than for the between-block designs. MAD and RMSE of the five-split within-block designs are 31% and 23% lower than those of the between-block designs. For the ten-split designs they are 41% and 40% lower than those of the between-block design. However, the MAD and RMSE of the within-block designs are comparable to those of the constrained between-block designs. The optimal within-block designs are shown in Figure 3.

[INSERT FIGURE 3 ABOUT HERE]

The estimates of the variances of the responses to the questions for the prior data, full and split questionnaires (after imputation) are shown in Table 5. As can be seen from the table, the prior estimates are close to complete questionnaire estimates of the current study. This illustrates the value of such prior estimates for the construction of split designs, but we further investigate the sensitivity of the optimal between- and within-block designs to these prior parameter values.

For this purpose, randomly draw 50 sets of values from the sampling distribution of the parameters obtained from the initialization data and obtain optimal ten-split unconstrained and constrained between-block designs and within-block designs based on each of these sets. On average, we found 9.7 splits to be the same across these replications for the unconstrained between-block design[6]. For the constrained ten-split between-block design we find a lower average number of corresponding splits, 5.5. For the within-block design, on average only 2.2 splits were the same. Clearly, the within-block design is much more sensitive to the choice of the prior than the between-block designs. The size of the full candidate split set, as well as the use of the greedy design generating algorithm contribute to the high prior sensitivity of the within-block design. We find the sensitivity of in particular the between-block design to the prior specification highly satisfactory.

[INSERT TABLE 5 ABOUT HERE]

FIELD STUDY

The above analysis illustrates that optimally designed split questionnaires can be beneficial, but only address that issue from a statistical perspective. In this section, we look into the behavioural issues of providing subjects split-questionnaires. We conducted a field experiment to investigate whether with split questionnaires one may reduce boredom, fatigue, and completion time, which ultimately should increase the quality of data. We will also investigate respondents' attitudes towards the questionnaires, and assess whether using split questionnaires improves the reliability of constructs, compared to the full questionnaire.

[6] The maximum is 10, if all prior values yield exactly the same design, since there are ten splits in the design.

For the field study, we use the exact questionnaire that was used in the empirical study above. We asked additional questions about boredom, which is scaled 1 (not at all bored) to 9 (extremely bored), fatigue which also is scaled 1 (not at all tired) to 9 (extremely tired). In addition, we assessed attitudes towards the questionnaire (three questions, five-point scale: repetitive-varied, very long-very short, boring-stimulating). We tested the full questionnaire, a ten-split between-block design, and a ten-split within-block design (see above) each on 63 subjects recruited from the subject pool from [withheld for confidentiality]. In total 189 subjects responded to 21 versions of the questionnaire that were displayed on computer screens in the experimental lab. Computer aided questionnaires allowed us to record the exact time it took respondents to complete them. These average times to complete the full and split questionnaires differed significantly, 8 minutes for the complete, and about 6 minutes for each of the split questionnaires. This is a significant reduction of about 25% in completion time, for a 50% reduction in the number of questions. Note that even the full questionnaire with 65 questions can be completed relatively quickly -the longest it took any respondent was 10 minutes-, which makes it more difficult to identify the behavioural effects of the split questionnaires.

The mean scores for the scales are shown in Table 6. A MANOVA across all measures reveals a significant difference between the complete and between-block design ($p<0.01$) and the complete and within-block design ($p<0.01$), but not the between the latter two. The mean boredom score for the full questionnaire is 5.44, which is significantly higher than that for the within-block questionnaire, which is 4.98. The differences with the between-block design, which has an intermediate boredom score of 5.23, are not significant. This may indicate that feelings of boredom are primarily caused by repetition of the relatively similar questions within blocks, which occurs less in the within-block design. The respondents that completed the full questionnaire report feeling more tired than those receiving the between-block design, the mean scores being 4.32 and 3.57. The within-block tiredness score is intermediate, 3.73, and not significantly different from the other two. This may point to feelings of tiredness being more strongly associated with switching between different topics, which occurs less often in the between-block design due to a reduction of the number blocks. The split questionnaire designs are evaluated more favorably than the complete questionnaire, the between- and within-block designs being seen as less repetitive (5.32 and 4.20 versus 5.68) and less boring (4.77 and 4.42 versus 4.94) than the complete questionnaire. The scores for the within-block design are significantly better than those for the between-block design. The within-block design is also considered to be significantly less long than the complete questionnaire design (3.13 versus 3.68; and 3.54 for the between-block design, which is not significantly different from the former two). The shorter perceived duration of the within-block design may be associated with its lower perceived boredom discussed above, since its actual duration is about 20 seconds longer than that of the between-block design (the longer duration may have to do with the reading and processing of the separate instructions for each block).

[INSERT TABLE 6 ABOUT HERE]

In short, split questionnaire designs decrease completion time, fatigue, boredom and non-response and are evaluated more positively by respondents, where it seems that the within-block design has a somewhat more favorable behavioral effect than the between-block design. These effects may impact the quality of the data. For each of the three questionnaires, respondents could skip every question displayed on the screen. There were 33 skip-responses for the full questionnaire, 7 for the between- and 5 for the within-block design. These responses start only after the first twelve questions and mostly occur in the last half of the questionnaires. This indicates that the use of split questionnaires may substantially reduce item non-response. Second, the effect of the questionnaire design on the average item variances and Cronbach's alpha were investigated. The questionnaire consists of 13 constructs that are each measured with several items. There were no statistically significant differences in the average Cronbach's alpha, estimated after multiple imputation of the missing data of the between- and within-block split questionnaire designs. But, we did find significant differences in item variances between the full and split-questionnaire designs. The differences between between-block and within-block designs are not significant. The average item variance for the full questionnaire is 3.34, which is significantly higher that for the between-block design, with 2.36, and the within-block design, 2.30. This means that subjects who answered the questions in the within-block or between-block design responded to the items that measure the same construct more consistently. Thus, the quality of the data we obtained from the between-and within-block split questionnaire designs tend to be better than that of the full questionnaire. Again, we note that with a maximum average completion time of eight minutes the complete questionnaire is relatively short. For longer questionnaires the effects may be even larger.

CONCLUSION

Split questionnaires present opportunities for application in consumer panels, offering the potential to obtain higher quality information from respondents faster and at a substantially lower cost. This paper first proposes a methodology to split questionnaires optimally into sub-components at minimal information loss, by applying optimal experimental design methods. We proposed the Kullback-Leibler Distance as a design criterion, applied the Modified Federov algorithm to search over the design space, and illustrated that good designs can be constructed rapidly in spite of the demanding task. Split questionnaire designs were shown to have desirable statistical and behavioral properties, relative to complete questionnaires or questionnaires constructed with ad-hoc methods.

We have investigated two different types of split questionnaire designs based on the contextual placement of questions in blocks. The first method, producing between-block designs, places blocks as a whole into different split-versions of the questionnaire. Optimizing the allocation of the blocks across the splits is a much more feasible task than allocating individual questions to splits. Additional constraints, such as on the number of blocks per split, can easily be accommodated and may further reduce the number of questions asked from each respondent. Between-block designs result in estimates close to those obtained from the complete data and reduce completion time and respondent fatigue. The second method, producing within-block designs, is based on choosing questions in each block. For these designs, the optimization task is very demanding, so that we needed to use a greedy algorithm to find the optimal design. As a consequence, the within-block designs are not strictly optimal nor can they easily be constructed to be fully identified. But, they do provide improved efficiency, yielding parameter estimates that are closer to the complete data estimates and less missing information than designs constructed with heuristic procedures. Their performance in terms of parameter estimates and missing information tends to be better than that of the between-block designs, but they are substantially more sensitive to the values of the prior estimates.

Our field study shows that the behavioral reaction of respondents to split questionnaires is more favorable than to the complete questionnaire, in terms of duration, boredom, and fatigue amongst others. The response to within-block designs tends to be more positive than that to a between-block design, since respondents feel less bored, and think that the questionnaire is less long, boring and repetitive. The between-block design, however, results in less respondent fatigue. The choice between the within-block and between-block designs may therefore be based on either statistical or behavioral criteria. From our investigation, it appears that the between-block design has better statistical properties, since it is feasible to construct fully identified designs with little sensitivity to the prior estimates. However, the within-block design still performs quite satisfactorily, yields parameter estimates comparable to constrained between-block designs, and elicits a more positive reaction form respondents. However, the high sensitivity of these designs to prior estimates warrants further study.

The validity of the prior knowledge to construct the split-questionnaire design is an important issue. Whereas prior knowledge can be easily obtained in panel or tracking surveys conducted on a regular basis with almost identical questions and blocks, it may be less easy to obtain in other settings. In those cases subjective prior distributions for the model parameters can be assessed, which in many cases would involve the elicitation of priors from consumers, decision makers or other subject-matter experts. Chaloner (1996) provides an overview of the various approaches to elicitation based on the ways people think about and update probabilistic statements. It is of interest to consider prior uncertainty on the parameters in constructing the designs, and to construct designs integrating the design criterion over the prior distribution of the parameters (Sandor and Wedel 2001). This may in particular be worthwhile for within-block designs, which were revealed to have high sensitivity to the prior specification. For between-block designs, in particular in panel data applications such as the one presented above, this may not be needed, since the prior parameter values can be fairly precisely estimated from the available pilot data, and the designs themselves were shown to be quite insensitive to the prior parameter values. We leave these issues for future research.

APPENDIX A: KL-DISTANCE FOR MIXED DATA

The incomplete data log-likelihood of mixed data is derived below using the general location model (Olkin and Tate1961; Krzanowski 1975). We have data matrix $Y_{N\times(p+q)}$=(X,Z), where X=(X_1,.., X_p)′ and Z=(Z_1,…,Z_q) represent the continuous and categorical variables, respectively. Each column variable in Z, z_j has c_j levels, and these categorical variables form a q-dimensional contingency table with a total number of cells $C = \prod_{j}^{q} c_j$. The frequencies in this table are contained in W = ($w_{c_1}, w_{c_2}, ... w_{c_q}$). The marginal distribution of the categorical variable Z is multinomial($w \mid \pi = (\pi_1, \pi_2, ... \pi_c)'$ ~multinomial (π) with $\sum_{i=1}^{C} \pi_i = 1$) and the conditional distribution of the continuous variables (X) given categorical variables (Z) (i.e. given a particular cell) is multivariate normal with different means across the cells defined by the categorical variables, but with a common covariance matrix ($X \mid Z = w, \mu_i, \Sigma \sim N(\mu_i, \Sigma)$, where μ_i is the mean of X in the cell specified by z, and Σ is the common conditional covariance of X across cells of the contingency table). The KL-distance in this case reduces to the incomplete-data log-likelihood:

$$L\ (\mu_z, \Sigma, \pi \mid D) = \sum_{i=1}^{N} \log f(x_{i,obs} \mid C_i, \mu_z, \Sigma) + \sum_{c\varepsilon_i}^{N} \log f(\pi_c)$$

(A1)
$$= -\frac{1}{2} \sum_{i=1}^{N} p_i \log(2\pi) - \frac{1}{2} \sum_{i=1}^{N} \log \left| \Sigma_{i,obs} \right|$$
$$+ \sum_{i=1}^{N} \log \left[\sum_{c \in C_i} \pi_i \exp \left\{ -\frac{1}{2} (X_{i,obs} - \mu_{i,obs,c})' \Sigma_{i,obs}^{-1} (X_{i,obs} - \mu_{i,obs,c}) \right\} \right] .$$

where $X_{i,obs} = X_{ij} d_{ij}$ where d_{ij} is the element of design matrix D.

REFERENCES

Adams, La Mar L. and Darwin Gale (1982), "Solving the Quandary Between Questionnaire Length and Response Rate in Educational Research," *Research in Higher Education*, 17 (3), 231-240.

Bar-Hen Avner and Daudin J.J. (1995), "Generalization of the Mahalanobis Distance in the Mixed Case", *Journal of Multivariate Analysis*, 53, 332-342.

Bean, Andrew G. and Michael J. Roszkowski (1995), "The Long and Short of It: When Does Questionnaire Length Affect Response Rate," *Marketing Research*, 7 (1), 21-26.

Berdie, Douglas R. (1989), "Reassessing the Value of High Response Rates to Mail Surveys," *Marketing Research*,1 (3), 52-64.

Chaloner, Kathryn (1996), "The Elicitation of Prior Distributions," to appear in *Bayesian Biostatistics*, eds. D. Berry and D. Stangl, New York: Marcel Dekker.

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

Dillman, Don A., Michael D. Sinclair, and Jon R. Clark (1993), "Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys," *Public Opinion Quarterly*, 57(3), 289-304.

Gelfand Alan E. and Smith Adrian F.M. (1990), "Sampling Based Approaches to Calculating Marginal Densities", *JASA*, Vol. 85, 398-409.

Gilula, Zvi, McCulloch, Robert E and Rossi, Peter E., (May 2004) "A Direct Approach to Data Fusion", Working paper, http://ssrn.com/abstract=549263

Good, Irving J. (1969), "Split Questionnaires I", The American Statistician, 23(4), 53-54.

_____________ (1970), "Split Questionnaires II", The American Statistician, 24(2), 36-37.

Heberline Thomas A. and Robert Baumgartner (1978), "Factors Affecting Response Rates to Mailed Questionnaires: a Quantitative Analysis of the Published Literature," *American Sociological Review*, 43(4), 447-462.

Herzog, Regula A. and Jerald G. Bachman (1981), “Effects of Questionnaire Length on Response Quality,” *Public Opinion Quarterly*, 45 (4), 489-504.

Huber Joel, Dick R. Wittink, John A. Fiedler, and Richard Miller (1993), “The Effectiveness of Alternate Preference Elicitation Procedures in Predicting Choice,” *Journal of Marketing Research,* 30 (1), 105-114.

Krzanowski Wojtek J. (1983), “Distance Between Populations Using Mixed Continuous and Categorical variables”, *Biometrika*, 70 (1), 235-243.

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

Kullback, Solomon and Leibler R.A. (1951), “On Information and Sufficiency,” *Annals of Mathematical Statistics*, 22 (1), 79-86.

Little, Roderick J.A. and Donald B. Rubin (1997), *Statistical Analysis with Missing Dat*a, New York: John Wiley & Sons

Novak Thomas P., Hoffman Donna L., and Yung Yiu-Fai (2000), “Measuring the Customer Experience in Online Environments: A Structural Modeling Approach”, *Marketing Science*, Vol. 19, No. 1, 22-42

O’ Hagan, Anthony (1994), Kendall’s Advanced Theory of Statistics, Volume 2B, *Bayesian Inference*, New York: Wiley & Sons.

Olkin, Ingram and Tate R.F. (1961), “Multivariate Correlation Models with Mixed Discrete and Continuous Variables”, *Annals of Mathematical Statistics*, 32 (2), 448-465

Orchard T. and Woodbury Max A. (1972), “A Missing Information Principle: Theory and Applications”, Proc. 6th Berkeley Symposium on Math. Statist. and Prob. 1, 697-715.

Raghunathan, Trivellore E. and James Grizzle (1995), “A Split Questionnaire Survey Design, ” *Journal of the American Statistical Association*, 90, 54-63.

Rassler Susanne (2002), Statistical Matching, A Frequentist Theory, Practical Applications, and Alternative Approaches, Springer.

Rodgers Willard L. (1984), “ An Evaluation of Statistical Matching”, *Journal of Business and Econometric Statistics*, 2,91-102.

Roszkowski, Michael J. and Andrew G. Bean (1990), “Believe It or Not! Longer Questionnaires Have Lower Response Rates,” *Journal of Business and Psychology*, 4(4), 495-509

Rubin, Donald B. (1987), *Multiple Imputation for Nonresponse in Surveys*, John Wiley&Sons.

Sándor, Zsolt and Michel Wedel (2001), “Designing Conjoint Choice Experiments Using Managers’Prior Beliefs”, *Journal of Marketing Research*, 38,430-449.

Shoemaker, David M. (1973), *Principles and Procedures of Multiple Matrix Sampling*, Cambridge, MA: Ballinger.

Sikkel, Dirk and Adriaan W. Hoogendoorn (1995), “Models for Monthly Penetrations with Incomplete Panel Data,” *Statistica Neerlandica*, 49 (3), 378-391.

Thayer, Dorothy T. (1983), “Maximum Likelihood Estimation of the Joint Covariance Matrix for Sections of Tests Given to Distinct Samples with Application to Test Equating”, *Psychometrika*, 48 (2), 293-297.

Table 1

CANDIDATE SPLIT SET FOR A FIVE BLOCK BETWEEN-BLOCK DESIGN

N_S	**Block 1 Q1-4**	**Block 2 Q5-8**	**Block 3 Q9-12**	**Block 4 Q13-16**	**Block 5 Q17-20**
1	0	0	0	0	0
2	1	0	0	0	0
3	0	1	0	0	0
4	0	0	1	0	0
5	0	0	0	1	0
6	0	0	0	0	1
7	1	1	0	0	0
8	1	0	1	0	0
9	0	1	1	0	0
10	1	1	1	0	0
11	1	0	0	1	0
12	0	1	0	1	0
13	1	1	0	1	0
14	0	0	1	1	0
15	1	0	1	1	0
16	0	1	1	1	0
17	1	1	1	1	0
18	1	0	0	0	1
19	0	1	0	0	1
20	1	1	0	0	1
21	0	0	1	0	1
22	1	0	1	0	1
23	0	1	1	0	1
24	1	1	1	0	1
25	0	0	0	1	1
26	1	0	0	1	1
27	0	1	0	1	1
28	1	1	0	1	1
29	0	0	1	1	1
30	1	0	1	1	1
31	0	1	1	1	1
32	1	1	1	1	1

Note: The size of the restricted split is 26 by excluding the splits with indices 1 to 6.

Table 2

PERFORMANCE OF THE MODIFIED-FEDEROV ALGORITHM

K	# of Possible designs (N_D)	Complete Enumeration (sec.)	Modified Federov Algorithm (sec.)[1]
5 splits	65780	260	20
10 splits	5311735	10456	50
15 splits	7726160	13343	78

[1] The modified Federov Algorithm results are based on 1000 random starts.

Table 3
SIMULATION RESULTS FOR BETWEEN-BLOCK DESIGNS

Design	Unconstrained 10 Splits		Constrained 10 Splits		Unconstrained 5 Splits		Constrained 5 Splits	
Measure	SQD[a]	MSD	SQD	MSD	SQD	MSD	SQD	MSD
MAD	3.143	3.730	2.471	2.773	2.817	3.210	3.001	3.102
RMSE	3.454	4.283	2.753	3.252	3.288	3.764	3.514	3.701
γ[b]	0.243	0.317	0.217	0.284	0.269	0.306	0.294	0.304
% Missing	0.600	0.600	0.500	0.500	0.533	0.533	0.500	0.500
BIC	5232	7193	8777	8989	4570	8170	8764	8796
logL(D)	-2608	-3589	-4380	-4486	-2277	-4077	-4374	-4390

[a] SQD = Optimal Split Questionnaire Design, MSD= Matrix Sampling Design

[b] γ is the fraction of missing information.

Table 4

COMPARISON OF DESIGNS ON EMPIRICAL DATA

BETWEEN-BLOCK DESIGNS								
	Unconst. 10 Splits		Const. 10 Splits-5Blocks/Split		Unconst. 5 Splits		Constr. 5 Splits-5 Blocks/Split	
	SQD	MSD	SQD	MSD	SQD	MSD	SQD	MSD
MAD	0.265	0.483	0.169	0.197	0.180	0.277	0.148	0.159
RMSE	0.378	0.682	0.240	0.319	0.262	0.399	0.203	0.215
γ	0.143	0.182	0.074	0.134	0.140	0.170	0.089	0.109
%Missing	0.735	0.735	0.492	0.492	0.662	0.662	0.440	0.440
BIC	18410	30298	57284	57655	15070	38740	64489	64675
logL(D)	-9195	-15139	-28631	-28817	-7525	-19360	-32234	-32327

WITHIN-BLOCK DESIGNS						
	10 splits			5 splits		
	SQD	RQD	PCA	SQD	RQD	PCA
MAD	0.156	0.163	0.164	0.125	0.125	0.129
RMSE	0.227	0.243	0.251	0.201	0.211	0.216
γ	0.078	0.087	0.084	0.056	0.060	0.058
%Missing	0.515	0.515	0.515	0.406	0.406	0.406
BIC	44134	45186	45085	54890	55126	54979
logL(D)	-22056	-22582	-22532	-27434	-27552	-27479

[a] SQD = optimal Split Questionnaire Design, MSD= Matrix Sampling Design, RQD = Random Questionnaire Design, PCA = Principal Components Design

Table 5

VARIANCE ESTIMATES AFTER IMPUTATION[1]

	Full Wave 1	Full Wave 2	Between Con. SQD	Between Unc. SQD	Within SQD
1	2.29	2.27	2.27	2.16	2.36
2	2.56	2.38	2.38	2.34	2.54
3	1.92	2.22	2.22	2.16	2.20
4	1.96	2.13	2.13	2.14	2.08
5	2.33	2.15	2.15	2.19	2.41
6	4.35	4.14	4.66	5.02	4.33
7	2.63	2.36	2.50	2.74	2.34
8	2.82	2.81	3.02	2.97	3.17
9	2.29	2.52	2.79	2.49	2.76
10	1.69	1.89	1.98	1.98	2.19
11	3.08	3.11	3.30	4.01	3.13
12	2.42	2.51	2.74	2.72	2.88
13	2.69	2.28	2.71	2.71	2.32
14	1.86	2.18	2.18	2.31	2.24
15	3.77	3.62	3.72	3.87	4.03
16	4.31	3.87	4.05	3.91	4.08
17	5.48	4.62	4.74	4.66	5.21
18	3.25	3.54	3.60	3.67	3.59
19	4.97	4.62	5.25	5.23	5.03
20	4.79	4.86	5.29	5.63	6.46
21	3.08	2.91	3.08	3.55	3.06
22	2.87	2.90	3.07	2.99	3.24
23	3.06	3.43	3.59	4.09	3.61
24	5.22	5.36	6.07	6.03	5.75
25	2.27	2.04	2.28	2.53	2.28
26	4.40	4.08	4.30	4.53	4.40
27	5.33	5.49	5.97	6.34	5.49
28	3.66	4.20	4.59	5.24	4.35
29	2.76	2.96	3.08	3.20	3.02
30	4.83	4.87	5.42	6.04	5.10
31	3.45	3.88	4.59	5.37	3.97
32	4.12	4.25	4.65	5.51	4.64
33	1.43	1.41	1.71	1.61	1.60
34	3.19	3.09	3.45	3.47	3.64
35	3.41	3.51	3.96	3.95	4.29
36	2.17	1.78	1.88	1.88	1.84
37	1.96	1.89	2.14	2.07	1.98
38	2.49	2.47	2.76	3.03	2.58
39	1.87	1.84	2.06	1.92	1.88
40	2.04	2.29	2.65	2.07	2.49
41	2.80	2.79	2.93	4.54	3.16
42	3.85	4.00	4.45	5.19	3.95
43	2.90	2.89	3.12	3.97	3.08
44	4.59	4.34	4.85	7.28	4.41
45	4.13	4.04	4.66	4.97	3.96
46	2.95	2.92	3.37	4.76	3.02
47	3.04	3.20	3.75	4.52	3.52
48	4.87	4.60	5.64	6.23	4.66
49	4.86	4.66	4.77	6.09	4.70
50	3.77	3.96	4.66	5.84	4.51
51	3.05	3.05	3.11	3.49	3.19
52	2.13	2.22	2.52	2.96	2.25
53	5.38	5.48	5.85	7.17	5.50
54	4.89	4.59	5.19	6.51	5.00
55	3.19	3.47	4.25	5.62	3.44
56	5.03	4.67	5.19	7.20	4.79
57	2.94	3.12	3.44	4.03	3.29
58	2.93	2.77	2.99	3.78	2.94
59	3.52	3.60	3.81	5.00	3.58
60	6.78	6.74	7.07	7.81	6.91
61	4.54	4.68	4.87	5.26	4.75
62	5.21	5.23	5.37	5.97	5.56
63	1.07	1.12	1.19	1.23	1.27
64	1.66	1.84	1.85	1.95	1.93
65	0.56	0.53	0.55	0.60	0.52

[1] From the full questionnaire from the first and second wave survey, the constrained and unconstrained between- and within ten-split optimal designs

Table 6

ITEM MEANS AND SD'S FROM THE FIELD EXPERIMENT

	FULL QUESTIONNAIRE	**BETWEEN-BLOCK SQD**	**WITHIN-BLOCK SQD**
Duration	476.92	344.48^{a1}	364.02^{b1}
	(95.01)	(146.552)	(93.57)
Boredom	5.44	5.23	4.98^{b1}
	(2.09)	(1.95)	(2.00)
Fatigue	4.32	3.57^{a2}	3.73
	(2.55)	(2.27)	(2.02)
Repetitive	5.68	5.32^{c1}	4.70^{b1c1}
	(1.37)	(1.22)	(1.78)
Long	3.68	3.54	3.13^{b1}
	(1.54)	(1.56)	(1.25)
Boring	4.94	4.77^{c1}	4.42^{b1c1}
	(1.28)	(1.11)	(1.25)
Cronbach's alpha	0.66	0.66	0.67
Item Variance	3.34	2.36^{a1}	2.30^{b1}

Notes: The values in parenthesis are standard deviations. N=189. Duration mean values are in seconds. Superscripts indicate the significance of the differences between means of the full & between- (^{a}), full & within- (^{b}) and between- & within- (^{c}) block questionnaires; ^{1} p=0.05 , ^{2} p=0.10

Figure 1

OPTIMAL UNCONSTRAINED BETWEEN-BLOCK DESIGNS

FOR THE EMPIRICAL DATA

THE OPTIMAL 10-SPLIT UNCONSTRAINED BETWEEN-BLOCK SQD

Resp.No.	**Block 1 Q1-5**	**Block 2 Q6-13**	**Block 3 Q14-18**	**Block 4 Q19-25**	**Block 5 Q26-31**	**Block 6 Q32-40**	**Block 7 Q41-50**	**Block 8 Q51-59**	**Block 9 Q60-65**
1-115									
116-230									
231-345									
346-460									
461-575									
576-690									
691-805									
806-920									
921-1035									
1036-1150									

THE OPTIMAL 5-SPLIT UNCONSTRAINED BETWEEN-BLOCK SQD

Resp.No.	**Block 1 Q1-5**	**Block 2 Q6-13**	**Block 3 Q14-18**	**Block 4 Q19-25**	**Block 5 Q26-31**	**Block 6 Q32-40**	**Block 7 Q41-50**	**Block 8 Q51-59**	**Block 9 Q60-65**
1-230									
231-460									
461-690									
691-920									
921-1150									

Note: shaded are observed, blank are missing blocks.
Note: Description of Blocks:
Block 1: Five questions about the role of the Web in life.
Block 2: Eight questions about the feeling while using the Web
Block 3: Five questions related to the Web activities feeling while using the Web
Block 4: Seven questions about and perceptions on using the Web
Block 5: Seven questions about attitudes and perceptions on using the Web
Block 6: Eight questions about people feelings towards using the Web
Block 7: Ten questions on attitudes and perceptions
Block 8: Nine questions about attitudes and perceptions on using the Web
Block 9: Six questions about flow and usage of the web.

Figure 2

THE OPTIMAL CONSTRAINED BETWEEN-BLOCK DESIGNS

FOR THE EMPIRICAL DATA

THE OPTIMAL 10-SPLIT 5-BLOCK BETWEEN-BLOCK SQD

Resp.No.	**Block 1 Q1-5**	**Block 2 Q6-13**	**Block 3 Q14-18**	**Block 4 Q19-25**	**Block 5 Q26-31**	**Block 6 Q32-40**	**Block 7 Q41-50**	**Block 8 Q51-59**	**Block 9 Q60-65**
1-115									
116-230									
231-345									
346-460									
461-575									
576-690									
691-805									
806-920									
921-1035									
1036-1150									

THE OPTIMAL 10-SPLIT 5-BLOCK BETWEEN-BLOCK SQD

Resp.No.	**Block 1 Q1-5**	**Block 2 Q6-13**	**Block 3 Q14-18**	**Block 4 Q19-25**	**Block 5 Q26-31**	**Block 6 Q32-40**	**Block 7 Q41-50**	**Block 8 Q51-59**	**Block 9 Q60-65**
1-230									
231-460									
461-690									
691-920									
921-1150									

Note: shaded are observed, blank are missing blocks.
Note: Description of Blocks:
Block 1: Five questions about the role of the Web in life.
Block 2: Eight questions about the feeling while using the Web
Block 3: Five questions related to the Web activities feeling while using the Web
Block 4: Seven questions about and perceptions on using the Web
Block 5: Seven questions about attitudes and perceptions on using the Web
Block 6: Eight questions about people feelings towards using the Web
Block 7: Ten questions on attitudes and perceptions
Block 8: Nine questions about attitudes and perceptions on using the Web
Block 9: Six questions about flow and usage of the web.

Figure 3

THE OPTIMAL WITHIN-BLOCK DESIGNS FOR THE EMPIRICAL DATA

THE OPTIMAL 10-SPLIT WITHIN-BLOCK SQD

Bl. 1 Q1-5	**Bl. 2 Q6-13**	**Bl. 3 Q14-18**	**Bl. 4 Q19-25**	**Bl. 5 Q26-31**	**Bl. 6 Q32-40**	**Bl. 7 Q41-50**	**Bl. 8 Q51-59**	**Bl. 9 Q60-65**
00110	00000101	00101	1100000	001101	000011111	0110001100	011111100	111010
11111	11111111	11111	1000100	1101010	010000010	0011100100	111000111	011111
00011	10000001	10100	1010000	0101010	000100001	1100110100	011011111	100010
10010	01000100	01001	0100001	0111110	110111110	0111110010	100111001	110010
10100	01010000	00111	1000001	1001001	101011100	1110111010	110011010	010011
01100	10010000	00011	0010010	1101110	100100010	0100011110	010111011	001101
00101	00101000	11000	1000010	1110101	000111100	1001111101	001011001	110101
11000	10010000	10010	0010001	1100110	010111110	1011011011	110011101	010001
01001	01000100	00110	0011000	0010011	111000011	1110010111	011000001	110111
10001	00110000	10001	0001001	0111100	101101111	0000000011	101011010	110011

THE OPTIMAL 5-SPLIT WITHIN-BLOCK SQD

Block 1 Q1-5	**Block 2 Q6-13**	**Block 3 Q14-18**	**Block 4 Q19-25**	**Block 5 Q26-31**	**Block 6 Q32-40**	**Block 7 Q41-50**	**Block 8 Q51-59**	**Block 9 Q60-65**
00110	01101101	00110	1010110	1111111	01100100	1110011101	000101010	101011
10010	10000101	11111	0101100	1101110	11110111	0000110011	011111101	011101
11111	00100011	01110	1011110	1101100	11011000	0111101011	111111111	100111
10100	01001111	00011	1101001	1001101	11101111	1101101110	100001101	110010
00101	11111111	10001	0110111	0101010	11001010	0001011001	000111000	001011