

Split Questionnaire Design for Massive Surveys

Feray Adigüzel¹, and Michel Wedel²

¹Erasmus University School of Economics, Rotterdam, The Netherlands

²University of Maryland, Robert H. Smith School of Business, College Park, USA.

Corresponding author: Feray Adiguzel, email: adiguzel@ese.eur.nl

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 using estimates from a first wave or pilot study. 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. We use Markov chain Monte Carlo procedures to impute missing values that result from the design. 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, incomplete block designs, and a heuristic procedure based on principal components analysis, using synthetic and empirical Web survey 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.

Keywords: Kullback Leibler distance, missing data, modified federov algorithm, multiple imputation, questionnaire design

1. Introduction

Researchers have traditionally collected data by means of often lengthy 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 major concern since it affects the quality of the data collected in several ways (Berdie 1989). Massive questionnaires take more time; induce the use of undesired response styles; increase respondent fatigue and boredom; and result in more nonresponse, item nonresponse, and early break-off (Adams and Gale 1982; Bean and Roszkowski 1995; Dillman 1991; Dillman, Sinclair, and Clark 1993; Heberline and Baumgartner 1978; Roszkowski and Bean 1990). Therefore, in practice, researchers frequently employ split questionnaires, that is dividing questionnaires into parts (or splits) and administering each part to a randomly selected group of respondents. However, this is very ad hoc and leads to a potential loss of information. There is a need for carefully designed split questionnaires which can help researchers collect information faster, cheaper, and more accurately.

We propose a method to design split-questionnaire surveys as an effective tool to reduce respondent burden, but with a minimal loss of information. 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. 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 (MSD) 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 efficiently split the questionnaire, such that the least information is lost. Raghunathan and Grizzle (1995) first proposed imputing the data that are missing due to (heuristically designed) split questionnaires using the Gibbs sampler. Their work is related to the literature on data fusion (Gilula, McCulloch, and Rossi 2006; Kamakura and Wedel 1997, 2000) and fits in the literature on multiple imputations of missing data (Little and Rubin 1997; Schafer 1997). There are two related aspects of the problem of splitting questionnaires: the design of the split questionnaire and the subsequent imputation of the missing values. Our focus is on the optimal design of the split questionnaires; standard Markov chain Monte Carlo (MCMC) methods can be used for the imputation of missing data due to that design (Raghunathan and Grizzle 1995). 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, to utilize the natural structure of the questionnaire, in which questions are placed in blocks, for example, one particular attitudinal or lifestyle trait are often administered in a single block. We use this block structure to generate a split questionnaire by either selecting entire blocks of questions (between-block design) or selecting questions from each block (within-block design). Design construction is easy and fast. Subsequently, we propose the use of the Kullback–Leibler distance (KLD) between the complete questionnaire and the split questionnaire data as an optimization criterion. We describe how to apply the modified Fedorov algorithm to find the optimal design among all possible designs. When the split questionnaire is administered, it results in data missing by design. Then, we impute the missing data with the Gibbs sampler (Raghunathan and Grizzle 1995; Schafer 1997). This imputation step is independent of the design step, so researchers can construct the SQDs without worrying about imputations. We compare the efficiency of our generated SQDs with MSDs (Shoemaker 1973) on synthetic data. Then, we apply our approach to data obtained from a Web questionnaire on attitudes toward and perceptions of using the Internet (Novak, Hoffman, and Yung 2000) and compare them with several alternative heuristic designs. Finally, in a field study, we investigate the extent to which the proposed SQD method may result in better data quality than the complete questionnaire.

2. Results

There are sixty-five questions in the Web Attitude and Perceptions Questionnaire and they are grouped into nine blocks according to content. Data are available for two waves of the study conducted in two consecutive years containing 500 and 1150 respondents, respectively. All data are complete. We also construct a field study with this questionnaire to indicate behavioral benefits. 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. 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.

2.1 Between-Block Designs

Table 1 reveals that for both the five and the ten-split between-block designs, the SQD improves substantially over the MSD, in terms of the KLD, the MAD (23%–70% lower), and the fraction of missing information (10%–32% lower). The results illustrate that the estimated parameters for the split design are close to the complete data estimates and that there is relatively little information loss given the percentage of questions eliminated. For this particular application, the five-split optimal split design results in a reduction of approximately 66.5% of the questions, with only a 35.8% information loss. With ten splits, we obtain a greater reduction in the number of questions, 73.7%, though the information loss is still only 38.9%.

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. 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.

2.2 Within-block Designs

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

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 within-block design is much more sensitive to the choice of the prior than the between-block designs. We find the sensitivity of in particular the between-block design to the prior specification highly satisfactory.

3. Field Study

For the field study, we use the exact questionnaire that was used in the empirical study above. We asked additional questions about boredom and fatigue. In addition, we assessed attitudes towards the questionnaire (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 each on 63 subjects recruited from the subject pool from. 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.

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 (Table 2). 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. 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 than 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.

4. Conclusions

Split questionnaires offer the potential to obtain higher-quality information from respondents at a faster rate and a lower cost. Split questionnaire designs were shown to

have desirable statistical properties relative to complete questionnaires or questionnaires constructed with ad-hoc methods. In addition, respondents reacted to a split questionnaire more favorably than to the complete questionnaire, in terms of completion time, boredom, and fatigue.

Between-block designs result in estimates close to those obtained from the complete data and reduce completion time and respondent fatigue. In within-block designs, questions within each block are placed into different splits. For these designs, the optimization task is demanding, and consequently the within-block designs are not strictly optimal and cannot be easily constructed to be fully identified. However, they provide improved efficiency and less missing information than designs constructed with heuristic procedures. Their statistical performance tends to be somewhat worse than that of the between-block designs, especially for small numbers of splits, and they are substantially more sensitive to the choice of priors, but they tend to solicit somewhat more favorable behavioral reactions and, in particular, lead to less boredom with the questions. The choice between the within-block and the between-block designs may be based on either statistical or behavioral criteria. From our investigation, it appears that the between-block design has better statistical properties because it is feasible to construct fully identified designs with less missing information and with little sensitivity to the prior estimates. Within-block designs are more difficult to construct and more sensitive to priors and lead to more fatigue and a longer time to complete. Thus, unless boredom is considered a problem, a between rather than within-design may be preferable.

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. We leave these issues for future research.

References

- Adams, L. and Gale D. (1982) "Solving the Quandary Between Questionnaire Length and Response Rate in Educational Research," *Research in Higher Education*, 17 (3), 231-240.
- Bean, A. G. and Roszkowski M. J. (1995) "The Long and Short of It: When Does Questionnaire Length Affect Response Rate," *Marketing Research*, 7 (1), 21-26.
- Berdie, D. R. (1989) "Reassessing the Value of High Response Rates to Mail Surveys," *Marketing Research*, 1 (3), 52-64.
- Dillman, D. A. (1991) "The Design and Administration of Mail Surveys," *Annual Review of Sociology*, 17, 225-249.
- Dillman, D.A., Sinclair M. D., and Clark J. R. (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.
- Gilula, Z., McCulloch, R. E and Rossi, P. (2006) "A Direct Approach to Data Fusion", *Journal of Marketing Research*, 43(1), 73-83
- Good, I. 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, R. A. and Bachman J. G. (1981) "Effects of Questionnaire Length on Response Quality," *Public Opinion Quarterly*, 45 (4), 489-504.

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

Little, R. J.A. and Rubin D. B. (1997) *Statistical Analysis with Missing Data*, New York: John Wiley & Sons

Novak, T.P., Hoffman D. L., and Yung Y. (2000) "Measuring the Customer Experience in Online Environments: A Structural Modeling Approach", *Marketing Science*, Vol. 19, No. 1, 22-42

Raghunathan, T. E. and Grizzle J. (1995) "A Split Questionnaire Survey Design," *Journal of the American Statistical Association*, 90, 54-63.

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

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

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

Sikkel, D. and Hoogendoorn A. W. (1995) "Models for Monthly Penetrations with Incomplete Panel Data," *Statistica Neerlandica*, 49 (3), 378-391.

Thayer, D. 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. Comparisons of Designs on empirical data on attitudes toward web usage

| <i>Between-Block Design</i> | | | | | | | | | | |
|-----------------------------|----------------------|------------|--------------------|------------|------------|----------------------|------------|--------------------|------------|------------|
| | <i>Ten Splits</i> | | | | | <i>Five Splits</i> | | | | |
| | <i>Unconstrained</i> | | <i>Constrained</i> | | | <i>Unconstrained</i> | | <i>Constrained</i> | | |
| | <i>SQD</i> | <i>MSD</i> | <i>SQD</i> | <i>MSD</i> | <i>IBD</i> | <i>SQD</i> | <i>MSD</i> | <i>SQD</i> | <i>MSD</i> | <i>IBD</i> |
| MAD | .798 | 2.631 | 1.172 | 1.431 | 1.731 | .887 | 1.145 | 1.311 | 1.447 | 1.510 |
| γ | .389 | .575 | .439 | .482 | .607 | .358 | .397 | .431 | .434 | .458 |
| % Missing | .737 | .737 | .494 | .494 | .532 | .665 | .665 | .449 | .449 | .440 |
| KLD | -6113 | -10,144 | -18,734 | -19,134 | -17,686 | -5051 | -12,666 | -20,688 | -20,873 | -21372 |

| <i>Within-Block Design</i> | | | | | | |
|----------------------------|---------------------------------|------------|------------|----------------------------------|------------|------------|
| | <i>Ten Splits Unconstrained</i> | | | <i>Five Splits Unconstrained</i> | | |
| | <i>SQD</i> | <i>MSD</i> | <i>PCA</i> | <i>SQD</i> | <i>MSD</i> | <i>PCA</i> |
| MAD | .731 | .794 | .945 | .994 | 1.106 | 1.292 |
| γ | .386 | .397 | .388 | .402 | .410 | .457 |
| % Missing | .626 | .626 | .626 | .508 | .508 | .508 |
| KLD | -14,372 | -14,749 | -14,695 | -18,738 | -19,176 | -18,784 |

Notes: This table shows a comparison of constrained (five blocks/split) and unconstrained five- and ten-split optimal designs with MSDs in terms of missing information, KLD, and MAD of (co)variances parameters.

Table 2. Results of the field experiment with 189 respondents

| | <i>Full Questionnaire</i> | <i>Between-Block SQD</i> | <i>Within-Block SQD</i> |
|------------|---------------------------|------------------------------|---------------------------------|
| Duration | 476.92 (95.01) | 344.48** (146.552) | 364.02 ^{ba} (93.57) |
| Boredom | 5.44 (2.09) | 5.23 (1.95) | 4.98 ^{ba} (2.00) |
| Fatigue | 4.32 (2.55) | 3.57*** (2.27) | 3.73 (2.02) |
| Repetitive | 5.68 (1.37) | 5.32 ^{ca} (1.22) | 4.70 ^{bacca} (1.78) |
| Long | 3.68 (1.54) | 3.54 (1.56) | 3.13 ^{ba} (1.25) |
| Boring | 4.94 (1.28) | 4.77 ^{ca} (1.11) | 4.42 ^{bcca} (1.25) |

*p = .05.
**p = .10.

Notes: This table presents item means (standard deviations) for the full questionnaire, the between-block questionnaire, and the within-block questionnaire. Duration mean values are in seconds. Superscripts indicate the significance of the differences between means of the full and within- (a), full and within- (b), and between- and within- (c) questionnaires from the multivariate analysis of variance.