Statistics Netherlands

The Leverhulme Trust

BADEN: Bayesian Adaptive Survey Design Network

State-of-the-Art

Statistics Netherlands (Centraal Bureau voor de Statistiek)

Barry Schouten

February 19th, 2015

1. Background information/introduction to your institution

Statistics Netherlands is the national statistical institute of the Netherlands. It produces social, economic and macro-economic indicators for the Dutch government and the European Union. It has roughly 2000 employees and its own data collection department. Indicators are based on primary and secondary data collection. For the primary data collection, it makes use of Web, Mail, CATI (computer-assisted telephone interviewing) and CAPI (computer-assisted personal interviewing) survey modes. For secondary data collection, the Dutch government installed a special legislation so that most administrative data can be linked temporarily to produce statistics and to support primary data collection. Statistics Netherlands maintains a Social Statistics Database in which all forms of data collection are linked and statistically edited.

Table 1: Overview of household surveys at Statistics Netherlands in 2015

Name of survey	Design	Under commission
Social Cohesion Survey	Web-CATI-CAPI	No
Experiences Survey	Web-CATI-CAPI	No
Survey of Consumer Sentiments	CATI	No
Labour Force Survey (LFS)	Web-CATI-CAPI	EU
NEA-BON	Web-PAPI	Yes
ZEA	Web	Yes
EU-Survey on Income and Living Conditions (SILC)	CATI	EU
Health Survey	Web-CAPI	No
Mobility Survey	Web-CATI-CAPI	Yes
Crime Victimisation Survey	Web-PAPI	Yes
ICT Survey (ICT)	Web-CATI	EU
Leisure Omnibus Survey	Web-CAPI	Yes
Survey on Health Care	Web-CATI-CAPI	Yes
Budget Survey	Web	No
Cultural Changes Survey	CAPI-PAPI	Yes
Housing Survey	Web-CATI-CAPI	Yes

In the BADEN network, we will focus on household survey data collection. Nevertheless, there is clear potential to use and translate theory and practical findings to business surveys. Table 1 gives an overview of the household surveys conducted in 2015, including the modes of data collection and their type in terms of financing.

The primary data collection for household surveys makes use of administrative data and paradata (data collection process data). Administrative data can be linked on an individual basis in order to define sampling strata, adaptive survey design strata and weighting/adjustment strata. During data collection linked data includes demographic data (gender, age, ethnicity, marital status, household), income data (personal and household), employment and allowance data, dwelling data (value, type of dwelling), regional data (urbanization at zip-code, municipality characteristics). Other administrative data become available with a time lag. Paradata comprises of call record data, interviewer observations, and Blaise audit trails and log files.

2. Current use and expertise of adaptive and/or responsive surveys in your institution

The Methodology department initiated research into adaptive survey design in 2008. Since 2008 a number of pilots has been conducted and various simulation studies have been performed. To date, no survey is fully based on the adaptive survey design paradigm, although a number of surveys does adapt data collection strategies to relevant subgroups, among which the Dutch Labour Force Survey. Since 2012, a number of internal projects has looked at the implementation of adaptive survey design. These projects advised positively about such an implementation. There were, however, three reasons why implementation was postponed: First, during the same period Statistics Netherlands redesigned virtually all surveys to multi-mode designs. Adaptive survey designs considering only nonresponse error, which is the original motivation for such designs, are too simple. It is known that measurement error plays an equally important role and cannot be ignored. The extension to measurement error makes theory more complex. Second, parallel to these redesigns, the current case management and survey administration systems became outdated and needs a thorough upgrade. Such an upgrade is complex given that data collection cannot be disturbed or paused. Adaptive survey design demands advanced case management, which would be partially manual under the current systems. Third, and most important for the BADEN objectives, response rates and mode population coverage rates showed considerable changes over the last three years; Web response rates showed a gradual decrease while telephone showed a steep decrease in coverage rates. As a result, the main input parameters to adaptive survey design are very unstable and require constant learning and updating. In January, Statistics Netherlands launched project Phoenix that has as primary goals the upgrade of survey case management, survey monitoring and survey analysis. It will include facilities to differentiate strategies and effort over relevant population subgroups. A debate is on-going currently, whether strata in such differentiation should be specific to the survey or should be chosen uniformly over surveys.

3. Current research on Adaptive Survey Designs in your institution

Research into adaptive survey design at Statistics Netherlands followed the work on indicators for representative response, termed R-indicators, within the 7th EU Framework project RISQ (www.risq-project.eu). Within RISQ a first successful pilot was conducted in 2009, linked to the Survey of Consumer Sentiments. Early research attempted to improve representativeness of survey response under constraints on the precision, budget, response rate and sample size. In 2009, a PhD was initiated which led to a dissertation in November 2013. The PhD, Melania Calinescu, started adaptive survey design research into nonresponse error, but gradually moved towards nonresponse and measurement error. The research used the Labour Force Survey (LFS) as case study, because this survey was undergoing a large mixed-mode redesign. For the LFS, an optimal design was derived which proposes a mix of mode strategies. This study led to a more in-depth follow-up research in which both data collection and social statistics department were involved. Within this follow-up the focus was on sensitivity and robustness.

Current research focusses on three topics:

1. Sensitivity and robustness of adaptive survey designs: Design input parameters are estimated in time series and are used to evaluate changes in optimal designs and robustness of optimal designs.
2. Learning and updating design parameters: A Bayesian approach towards both the estimation of design parameters and the optimization of strategy allocation is investigated. This research is closely linked to the first topic and falls largely within BADEN. Prior and posterior distributions are derived for response propensities, cost parameters and mode effects in three surveys that will be redesigned in 2015-2018. Tools are developed to compute the posteriors during data collection. Cost rules for optimization are investigated.
3. Legitimizing adaptive survey design: Adaptive survey design are criticised for adjusting imbalance or lack of representativeness that can be adjusted for afterwards in the estimation. Theory is developed to give conditions under which the designs are effective, even after adjustment, as well as empirical evidence.

4. Proposed research agenda for Bayesian Adaptive Survey Designs

A postdoctoral researcher will be contracted April 1st for a period of a year. Continuation of the position will be decided in the autumn based on progress, results and available budget. Apart from the postdoc, three researchers at the methodology department will be involved in research into adaptive survey design, and more specifically, a Bayesian approach towards such design. These researchers are Joep Burger, Nino Mushkudiani and Barry Schouten.

The agenda for the coming year is:

- Master thesis sensitivity analyses into adaptive survey design linked to the Labour Force Survey (LFS) and Housing Survey (OViN), June 2015;
- Invited paper session Adaptive survey design, WSC ISI, July 2015;
- Start of pilot linked to Consumer of Sentiments Survey, April 2015;
- Postdoc research into prior/posterior distributions, tools for monitoring, optimization under Bayesian distributions, April 2015 – April 2016;
- Research master Bayesian adaptive survey design, assistance in deriving priors/posteriors, construction of time series, September 2015 – June 2016;
- Additional research at methodology department depending on agenda BADEN following the kick-off meeting in February, March 2015 and further;

5. Collaborations planned

The methodology department collaborates with Utrecht University and VU University Amsterdam. Currently, a research master student from Utrecht University is investigating the sensitivity of adaptive survey designs to inaccuracy of design input parameters. A research proposal was submitted to Utrecht University for a follow-up to this project. A postdoctoral researcher will be contracted to support research into Bayesian adaptive survey design. This postdoc will have an affiliation with VU university Amsterdam.

6. References

- Schouten, B., Shlomo, N. (2015), Selecting adaptive survey design strata with partial R-indicators, Discussion paper 2015xx, Statistics Netherlands, The Hague, forthcoming.
- Calinescu, M., Schouten, B. (2015), Adaptive survey designs to minimize mode effects. A case study on the Dutch Labour Force Survey, Forthcoming in Survey Methodology

- Schouten, B., Calinescu, M., Burger, J. (2014), Adaptive mixed-mode survey designs accounting for mode effects – A case study on the Dutch Labor Force survey, Invited paper, Proceedings of Joint Statistical Meetings, August 3 – 7, Boston, USA
- Ouwehand, P., Schouten, B. (2014), Measuring representativeness of short term business statistics, Journal of Official Statistics, 30 (4), 623 – 649.
- Schouten, B., Cobben, F., Lundquist, P., Wagner, J. (2014), Theoretical and empirical evidence for balancing of survey response by design?, Discussion paper 201415, Statistics Netherlands, The hague, available at www.cbs.nl.
- Calinescu, M., Bhulai, S., Schouten, B. (2013), Optimal resource allocation in survey designs, European Journal of Operations Research, 226 (1), 115-121.
- Schouten, B., Calinescu, M., Luiten, A. (2013), Optimizing quality of response through adaptive survey designs, Survey Methodology, 39 (1), 29 – 58.
- Schouten, B., Calinescu, M. (2013), Paradata as input to monitoring representativeness and measurement profiles. A case study on the Labour Force Survey, pages 233 – 259, *In Improving Surveys with Paradata. Analytic Uses of Process Information*, ed. F. Kreuter.
- Luiten, A., Schouten, B. (2013), Adaptive fieldwork design to increase representative household survey response. A pilot study in the Survey of Consumer Satisfaction, Journal of Royal Statistical Society, Series A, 176 (1), 169 – 190.
- Schouten, J.G., Bethlehem, J., Beulens, K., Kleven, Ø., Loosveldt, G., Rutar, K., Shlomo, N., Skinner, C. (2012), Evaluating, comparing, monitoring and improving representativeness of survey response through R-indicators and partial R-indicators, International Statistical Review, 80 (3), 382 – 399.
- Schouten, J.G., Shlomo, N., Skinner, C. (2011), Indicators for monitoring and improving representativeness of response, Journal of Official Statistics, 27(2), 231 – 253.
- Schouten, J.G., Cobben, F., Bethlehem, J. (2009), Indicators for the representativeness of survey response, Survey Methodology, 35 (1), 101 – 113.