Statistics Netherlands

The Leverhulme Trust

BADEN: Bayesian Adaptive Survey Design Network

State-of-the-Art

University of Southampton

Gabriele Durrant

February 24th, 2015

1. Background information/introduction to your institution

The University of Southampton is based in the United Kingdom. The University hosts a number of departments and research institutes with interests in statistical methods, including the Department of Social Statistics and Demography, the Southampton Statistical Sciences Research Institute, the ESRC-funded National Centre for Research Methods (NCRM) and the ESRC-funded Administrative Data Research Centre for England (ADRCE).

A new UK wide network, the Administrative Data Research Network (ADRN) has recently been established driving forward the use and linkage of administrative data. The ADRCE is part of this network and works closely together with the UK Office for National Statistics (ONS) as part of this grant. The ADRCE and the ADRN is a multimillion investment into the improvement of data structure and data access in the UK. It has been established for 5 years initially (currently until 2018). As part of these activities the University of Southampton has recently created a new secure data laboratory to enable researchers from across the UK (and from abroad) to come to the University and to work on highly sensitive data, in particular linked administrative data. Currently the first project applications aiming to link administrative data sources to each other or linking survey data to administrative data are under way. It is difficult to say at this stage when exactly work of this nature can start due to difficulties in regulations and overcoming legal and political constraints. The secure data laboratory is already set up and fully functioning and usable.

The University of Southampton does not have its own data collection department. We use secondary data from other institutions. The University of Southampton works closely with the Office for National Statistics and other national statistical institutes (e.g. with Statistics Norway via Li-Chun Zhang). We also use data from national data archives (e.g. UK Data Archive at the University of Essex).

The National Centre for Research Methods is another multi-million investment over a 5 year period (currently until 2019). A range of UK academics are involved from three institutions, the University of Southampton, the University of Manchester (including Prof Natalie Shlomo leading the BADEN project) and the University of Edinburgh. Amongst a range of research projects (the one relevant to the BADEN project will be described below) the NCRM has a commitment to training in advanced quantitative methods, including short courses and online resources. Possibilities exists for the NCRM to support training activities in the area of paradata, adaptive and responsive survey designs, nonresponse and measurement error in sample surveys, multi-mode surveys, data linkage, use of

administrative data and other related topics (e.g. short courses and online resource module development).

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

The main people working on paradata and adaptive/responsive survey designs are Gabriele Durrant, a number of her PhD students and research fellows (e.g. Olga Maslovskaya, Jamie Moore), a lecturer (Solange Correa) and to some extent Prof Peter Smith and Prof Patrick Sturgis. Some of the work developed, although sometimes not currently implemented in adaptive and responsive survey designs, have wide ranging practical implementations and could be implemented in the future by a survey organisation.

The UoS focusses on the development of theories, strategies and methods as well as on implementation and applied work using secondary data analysis. It cannot implement adaptive and responsive survey designs directly since we do not run our own surveys. However, we work closely with survey organisations. This also includes Ipsos-Mori and TNS-BMRB in London (Prof Patrick Sturgis has close contacts). We hope that the BADEN project will facilitate further collaboration possibilities.

3. Recent and current research on Adaptive Survey Designs in your institution

From 2011 until August 2015 the University of Southampton run an ESRC-funded research project titled 'Use of Paradata (Field Process Data) in Cross-Sectional and Longitudinal Surveys' (principle investigator Gabriele Durrant). The project comprises three subprojects with the specific aims of: 1) investigating the use of call record data and interviewer observations to study nonresponse in longitudinal surveys, 2) providing insights into the effects of interviewing strategies and other interviewer attributes on response in longitudinal surveys, and 3) gaining knowledge about the measurement error properties of paradata. The work on paradata has direct implications for adaptive and responsive survey designs. This project on paradata is closely related to an earlier ESRC-funded project on 'Hierarchical Analysis of Unit-Nonresponse in Sample Surveys'.

Both ESRC funded projects organised an international research symposium on paradata (the second one was on 26th June 2014 in London at the Royal Statistical Society titled 'Paradata: From Survey Research to Practice' and the first one was in Dec 2010 also in London at the RSS, 'Recent Advances in the Use of Paradata in Social Survey Research').

The ESRC grant on paradata used primarily the UK 2001 Census nonresponse link study and Understanding Society, a large-scale longitudinal household survey in the UK, comprising relatively rich paradata (call record data and interviewer observations) and some information on interviewers. Since it is a longitudinal survey information on individuals and households is available from the first wave or prior waves. There are plans by the survey organisation (University of Essex) to link in the future respondents' information to external sources such as from administrative data (but only possible for cases that responded to the survey and consented to data linkage). Our work so far focussed on using sequence analysis to analyse call sequences, the development of models predicting length of call sequence and final outcome of call sequences, aiming to answer questions about how best to incorporate current and prior call history information and substantive information on cases (e.g. from administrative data and prior wave information), assessing nonresponse bias during field work (using call record data) and the use of discrete time event history models to investigates best times of contact and cooperation (as well as other correlates to establish contact and cooperation at the next call, i.e. during fieldwork) (see list of references below). Also a body of work has been developed analysing interviewer effects (on nonresponse rates, nonresponse bias and measurement error).

A closely related research project is the workpackage 1 of the ESRC National Centre for Research Methods, titled 'Analysing hierarchical time-dependent paradata to reduce nonresponse in large-scale surveys', where analysis techniques will be developed and applied in the substantive context of

using survey paradata to model unit nonresponse. The official start of the project is Oct 2015, although some work has already started. The project makes use of the UK 2011 Census nonresponse link study, a potentially very rich dataset for the UK context. The dataset comprises the outcome of several UK household surveys (indicator if person and/or household responded and type of response and nonresponse; probably usable 4 surveys), linked to variables from the UK 2011 Census at both the household and individual level, call record data and interviewer observations, and interviewer id. (Unfortunately, at least currently, the actual survey variables are not included, information on interviewers is not included). The work that has started on the dataset is analysing nonresponse bias across calls (ie. during fieldwork).

Recent and current research has focused on household surveys. In the BADEN network, we will continue to focus on household survey data. Nevertheless, there is clear potential to use and translate theory and practical findings to business surveys.

4. Proposed research agenda for Bayesian Adaptive Survey Designs

Currently two postdoctoral researchers are employed working on topics related to paradata and adaptive/responsive survey designs (Olga Maslovskaya - currently on maternity leave until Sept 2015, then on a three-year contract - and Jamie Moore, currently on a 4-year contract). A PhD student (funded by the National Centre for Research Methods and the ESRC) will start in Sept 2015 and will be working on a topic related to the NCRM workpackage 1. Potential topics are the exploration of linked datasources (paradata and administrative data) and interviewer effects in longitudinal surveys. Gabriele Durrant, Peter Smith and Patrick Sturgis have part of their time funded under the NCRM and the ADRCE grant and as such can work on topics related to the grants and the BADEN project.

Possibilities exists for the NCRM to support research and training activities in the area of paradata and adaptive and responsive survey designs (further topics may include: nonresponse and measurement error in sample surveys, multi-mode surveys, data linkage, use of administrative data and other related topics. NCRM could for example support a research event, one or several short courses and online resource module development.

The agenda for the coming year is:

<u>Conference activities in 2015:</u>

- Special contributed paper session on ‘The use of paradata (survey process data) in response analysis, response monitoring and responsive design’, ISI, July 2015.
    - Durrant, G.B., Maslovskaya, O., and Smith, P.W. (2015) Modelling Final Outcome and Length of Call to Improve Efficiency in Call Scheduling, to be presented at a special topic contributed session at the bi-annual conference of the International Statistical Institute, on ‘The use of paradata (survey process data) in response analysis, response monitoring and responsive design’, Rio, Brazil, August 2015.
    - Correa, S., Durrant, G.B. and P.W. Smith (2015) Assessing Nonresponse Bias using Call Record Data with Applications to a Longitudinal Study, to be presented at a special topic contributed session at the bi-annual conference of the International Statistical Institute, on ‘The use of paradata (survey process data) in response analysis, response monitoring and responsive design’, Rio, Brazil, August 2015.
- Discussant at the annual conference of the American Association for Public Opinion Research (AAPOR), session on: ‘Using Paradata During Data Collection and in Data Analysis: New Metrics to Address Perennial Problems’, May 2015, Florida.
- Co-organiser of the session ‘Modelling unit nonresponse and attrition processes’ at the ESRA conference (European Survey Research Association), with Carina Cornesse and Dr. Annelies Blom from the University of Mannheim, 13-17 July, Reykjavik, Iceland.
- Further presentations planned for 2015:

- Moore, J. Durrant, G.B, Smith, P.W. (2015) Are trajectories of dataset representativeness during survey data collection generalisable? Evidence from the 2011 Census Non-Response Link Study contributed paper prepared for the ISI 2015, Rio, (to be confirmed).
- Barbosa, D., Durrant, G.B. and P.W. Smith (2015) Interviewer Effects on Measurement Error, contributed paper to be presented at the bi-annual conference of the International Statistical Institute, Rio, August 2015 (to be confirmed).

Research work planned:

- Finalising the first paper on using the 2011 UK Census Nonresponse Link study (with Jamie Moore and Peter Smith) for submission to a journal.
- Writing up of a paper using data from wave 1 and 2 of Understanding Society (planned by Summer 2015), on how best to specify models to predict length of call sequence and call outcome conditioning on current and prior call history and prior wave information (work with Olga Maslovskaya and Peter Smith).
- Writing up of a book chapter: Using Quantitative Paradata to Improve the Data Quality in Social Surveys, invited paper, in Edwards, R., Phoenix, A., O'Connor, H. and Goodwin, J. (eds.) (2015) Working With Paradata, Marginalia and Fieldnotes: The Centrality of By-Products of Social Research.
- Defining the work programme for Olga Maslovskaya (to start again in Oct 2015), after her return from maternity leave this may include initially finalising already started work or working on any revisions of papers already submitted.
- Continuation and extension of joint work with Solange Correa and Peter Smith (e.g. use of imputation during data collection to improve estimation).
- Exploration of further work on case prioritisation using response scoring with Li-Chun Zhang.
- Additional research depending on agenda BADEN following the kick-off meeting in February and further.

5. Collaborations planned

The University of Southampton (UoS) collaborates with Statistics Norway (via Li-Chun Zhang who is employed both at Statistics Norway and is Professor at UoS). Southampton already plans to work closely with Natalie Shlomo from the University of Manchester (and leader of the BADEN project) as part of the NCRM research grant. As part of the ADRCE grant (and the Southampton-ONS methodology contract) Southampton works already closely with the Office for National Statistics (although there are also obstacles and hurdles). Dr Paul Smith who used to work for ONS for many years is now associate professor at UoS. Gabi has worked with Frauke Kreuter as part of the ESRC grant on paradata and further collaborations may emerge.

6. References

Correa, S., Durrant, G.B. and Smith, P.W. (2015) Assessing Nonresponse Bias using Call Record Data with Applications to a Longitudinal Study, submitted.

Durrant, G. B. and Kreuter, F. (Guest Editors) (2013): The Use of Paradata (Process Data) in Social Survey Research, *Special Issue on paradata, Journal of the Royal Statistical Society, Series A*, 176, 1.

Durrant, G.B. and D'Arrigo, J. and Müller, G. (2013) Modeling Call Record Data: Examples from Cross-Sectional and Longitudinal Surveys, in: Kreuter, F. (ed) (2013) *Improving Surveys with Paradata*, Wiley and Sons.

Durrant, G.B. and Kreuter, F. (2013) Editorial: The Use of Paradata in Social Survey Research, Journal of the Royal Statistical Society, Series A, Special issue: The Use of Paradata in Social Survey Research, 176, 1, 1-3.

Durrant, G.B., D'Arrigo, J. and Steele, F. (2011): Using Field Process Data to Predict Best Times of Contact Conditioning on Household and Interviewer Influences, Journal of the Royal Statistical Society, Series A, 174, 4, 1029-1049.

Durrant, G.B., D'Arrigo, J. and Steele, F. (2013): Analysing Interviewer Call Record Data by Using a Multilevel Discrete-Time Event History Modelling Approach, Journal of the Royal Statistical Society, Series A, Special issue: The Use of Paradata in Social Survey Research 176, 1, 251-269.

Durrant, G.B., Maslovskaya, O., Smith, P.W. (2015) Modelling Final Outcome and Length of Call to Improve Efficiency in Call Scheduling, conditionally accepted with *Journal of Survey Statistics and Methodology*, subject to very minor corrections.

Durrant, G.B., Maslovskaya, O., Smith, P.W. (2015) Sequence Analysis: A Graphical Tool for Investigating Call Record Data, to be submitted shortly.

Sinibaldi, J., Durrant, G.B. and Kreuter, F. (2013): Evaluating the Measurement Error of Interviewer Observed Paradata, Public Opinion Quarterly, Special issue: Topics in Survey Measurement and Public Opinion, 77, 1, 173-193.

In writing up stage:

Durrant, G.B., Maslovskaya, O., Smith, P.W. (2015) Modelling Final Outcome and Length of Call in a Longitudinal Study Taking Account of Prior Call Information, (in writing up stage).

Durrant, G.B. (2015) Using Quantitative Paradata to Improve the Data Quality in Social Surveys, invited paper, in Edwards, R., Phoenix, A., O'Connor, H. and Goodwin, J. (eds.) (2015) Working With Paradata, Marginalia and Fieldnotes: The Centrality of By-Products of Social Research, Edward Elgar, London, Ch. 3. (available in Summer 2015).

Sturgis, P., Brunton-Smith, I. and Williams, J. (2015) Fieldwork effort, response rate and the distribution of survey outcomes: a multilevel meta-analysis.

Barbosa, D, Durrant, G.B. and Smith, P.W.F (2015) Interviewer effects on nonresponse bias, in writing up stage.

Barbosa, D, Durrant, G.B. and Smith, P.W.F (2015) Interviewer effects on measurement error, in writing up stage.

Indirectly related:

Durrant, G.B. and D'Arrigo, J. (2014) Doorstep Interactions and Interviewer Effects on the Process Leading to Cooperation or Refusal, *Sociological Methods and Research*, 43, 490-518.

Durrant, G.B. and Steele, F. (2009): Multilevel Modelling of Refusal and Noncontact Nonresponse in Household Surveys: Evidence from Six UK Government Surveys, Journal of the Royal Statistical Society, Series A, 172, 2, 361-381.

Durrant, G.B., Groves, G., Staetsky L. and Steele, F. (2010): Effects of Interviewer Attitudes and Behaviours on Refusal in Household Surveys. Public Opinion Quarterly, 74, 1, 1-36.

Steele, F. and Durrant, G. (2011): Alternative Approaches to Multilevel Modelling of Survey Noncontact and Refusal, International Statistical Review, 79, 1, 70-91.

Vassallo, R, Durrant, G.B., Smith, P. and Goldstein, H. (2015): Interviewer Effects on Nonresponse Propensity in Longitudinal Surveys: A Multilevel Modelling Approach, Journal of the Royal Statistical Society, Series A, 178, 1, 83-99.

Vassallo, R., Durrant, G.B. and Smith, P. (2014) Assessment of Multiple Membership Multilevel Models: An application to interviewer effects on nonresponse, submitted (revise and resubmit).

Vassallo, R., Durrant, G.B. and Smith, P. (2014) Separating Interviewer and Area Effects Using a Cross-Classified Multilevel Logistic Model: Implications for Survey Designs, under review.