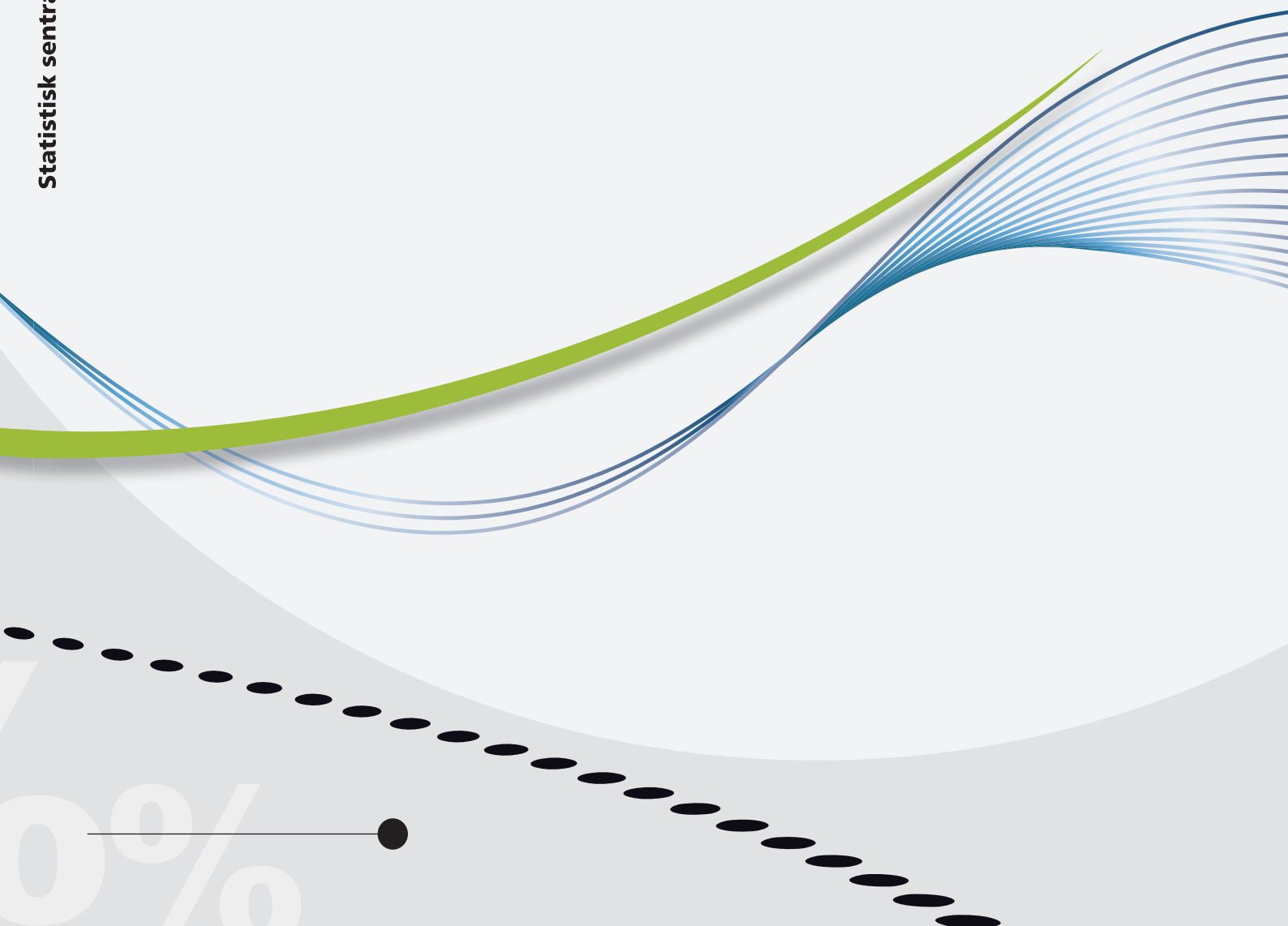


Susie Jentoft and Tora C. Löfgren

Response Propensities in the Norwegian Labour Force Survey

The move towards a responsive design



Susie Jentoft and Tora C. Löfgren

**Response Propensities in the Norwegian
Labour Force Survey**

The move towards a responsive design

Documents: In this series, documentation, method descriptions, model descriptions and standards are published

© Statistics Norway

When using material from this publication, Statistics Norway shall be quoted as the source.

Published 6 November 2015

ISBN 978-82-537-9247-7 (electronic)

Symbols in tables	Symbol
Category not applicable	.
Data not available	..
Data not yet available	...
Not for publication	:
Nil	-
Less than 0.5 of unit employed	0
Less than 0.05 of unit employed	0.0
Provisional or preliminary figure	*
Break in the homogeneity of a vertical series	—
Break in the homogeneity of a horizontal series	
Decimal punctuation mark	.

Preface

Collaboration on this paper started in 2014 as part of a European Statistical System (ESS) project on data collection methods for social surveys using multi modes (DCSS). It is an important area for Statistics Norway to research given the continual advances in technology and importance of how we communicate with survey participants. Recent years have shown a decrease in response rates in many social surveys creating a need for further investigation on the driving forces behind this trend. Understanding the mechanisms can allow us to adapt our data collection methods to maximize participation and allow us to continue producing high quality estimates for the country.

The analyses in this paper were performed by the authors with contributions from Trond Båshus from the Division for social surveys. Arnhild Torsteinsen, Bengt Oscar Lagerstrøm, Mathias Killengreen Revold (all Division for social surveys) and Dag Gravem (Division for methods) have all contributed to our understanding and interpretation of the results.

Statistisk sentralbyrå, 26. oktober 2015

Anne Sundvoll

Abstract

With response rates decreasing in the Norwegian Labour Force Survey (LFS) it is important for Statistics Norway to investigate ways to optimize response within a constraint budget. This paper addresses optimal times and days for respondents and looks at how response propensities and response rates change throughout the data collection period. Additionally we observe the progression of R-indicators and a key variable: employment. The data comes from quarters 2, 3 and 4 in 2013, including over 66 000 cases.

The average (mean) case length was around 5 contact attempts while the median was only 2 indicating a skew in the case length data; a few people are followed-up many times. We observed that response propensities show a bi-modal distribution and drop during the data collection period; those that have not been able to be contacted are increasingly harder to reach. The last timeslot (19:00 to 21:30) appears to be the best time for initial contact attempts in order to achieve an interview.

Our study provides some basic ground work on when an appropriate time to swap to a new mode may be. Two of the ways we may consider the best point to change are in terms of phase capacity and for maximizing survey representativity. We see that response rates are not greatly improved once 15 contact attempts have been made. However, our key variable: employment, and the R-indicators and partial R-indicators stabilise before this; already after around 5 contact attempts. We believe this gives a good indication that a switch to a cheaper follow-up mode may be warranted already at this point.

One option could be to base the timing of the swap on a responsive design, whereby certain groups swap to the final cheaper mode when the response rates within the group are above a certain level. Similarly, groups could be routed to the more expensive (but more efficient at recruiting subjects to the survey) telephone follow-up sooner, if initial response rates within the group are below a certain level. Simple control charts for this could be developed with basis in our figures in chapter 6, preferably where the response or outcome influence on the estimate is shown. This since we through this basic analysis have shown that some background variables get more skewed after a certain number of contact attempts.

We would also like to stress that a measure or treatment in responsive design is not only about offering the respondent or groups of respondents different modes; it may well consist of e.g. different contact strategies, differentiated information about the survey and use of incentives.

We suggest taking the necessary steps to implement an adaptive design/responsive design for the Norwegian LFS in the very near future to make the best use of resources. Additionally, a future experiment on the initial day and time of first contact would allow us to investigate how to optimise response rates, and improve resource allocation.

Contents

- Preface 3**
- Abstract..... 4**
- Contents..... 5**
- 1. Introduction 6**
- 2. Responsive design – a case of adaptive survey design..... 6**
- 3. Data structure..... 7**
- 4. Case length..... 7**
- 5. Response propensities..... 9**
- 6. Investigation of timing for first contact..... 12**
 - 6.1. First contact attempt for first wave of subjects 12
 - 6.2. Best day for contact..... 12
 - 5.3 Second contact attempt for first wave subjects 13
- 7. Phase capacity and mode switch in the Norwegian LFS..... 13**
 - 7.1. Representativity and R-indicators..... 14
 - 7.2. Our model 14
- 8. Concluding remarks and further analysis..... 17**
 - 8.1. Suggestions for a responsive design..... 17
 - 8.2. Further analysis..... 18
- References 19**
- Appendix A..... 20**
- Appendix B..... 21**

1. Introduction

Non-response in the Norwegian Labour Force Survey (LFS) has risen from 13 per cent in 2008 to around 20 per cent in 2012. With increased data collection costs it becomes more and more important to monitor the data collection and survey productivity and at the same time try to maintain data quality within a reasonable budget. This especially for interviewer administered modes where scarce resources can be used somewhere else in the production process. A responsive design could provide operations managers with the information they need to make more efficient use of their resources in real-time during the ongoing data collection.

To understand what costs in a data collection using interviewers one must know where most of the effort is spent. In general, the effort-demanding process can be divided into three parts; 1) making a contact with the person or household 2) getting cooperation from the person and 3) conducting the interview with the respondent (Mohl and Laflamme, 2007), and in our analysis, we concentrate on the first two stages.

This analysis uses paradata from the Norwegian LFS to investigate response propensities during the data collection period and how their distributions change after each contact attempt. This allows us a predictive way to identify problem groups while the collection period is still running. By additionally investigating the timing of contacts in relation to these problem groups we can identify ways to maximize their response propensities. The analysis provides groundwork for developing a responsive data collection design for the Norwegian LFS, where available information will be used continuously to modify the data collection for remaining cases. All figures and analyses were done using the software R 3.1.2 (R Core Team, 2014).

Areas of study for this paper are:

1. Which days and times are best for the highest response rates in general.
2. At what point is it not worth continuing to ring someone based on their history, i.e. when have we reached phase capacity such as we should swap to a different mode, e.g. web.

2. Responsive design – a case of adaptive survey design

In adaptive survey design different persons or households receive different treatments. In terms of data collection, treatment often refers to different modes, but it can also be different information setups for different respondent groups or different ways of sending reminders. The modes can be defined before the field period starts but strategies may also change due to what is observed during the data collection. Adaptive survey designs originate from medical statistics and clinical trials. Responsive design is a special case of adaptive design proposed by Groves and Heeringa (2006). They point out that responsive survey designs should:

- pre-identify a set of design features potentially affecting costs and errors of survey estimates
- identify a set of indicators of the cost and error properties of those features and monitor those indicators in initial phases of data collection
- alter the features of the survey in subsequent phases based on cost-error trade-off decision rules
- combine data from separate design phases into a single estimator

So, responsive designs are organized about design phases, use indicators of potential survey error and process data to determine when design changes should be made on an ongoing basis during data collection (Wagner et. al., 2012; Groves and Heeringa, 2006). A design phase is a time period of the data collection where the same sample design, recruitment protocol, mode of collection etc. is followed. A survey may for instance start with a web questionnaire, followed by a telephone phase, and then finally a swap to a face-to-face interview phase before data collection is finished.

For monitoring and managing the data collection a set of indicators has to be developed. To be operationally useful, the indicators for survey performance should have some of the characteristics listed below (Laflamme, 2009):

- they should be easily understood and interpreted
- they have to be measurable and comparable at any point during data collection
- they should always be updated during data collection, this to reflect changes
- they should be relevant, interpretable and comparable at different aggregation levels

3. Data structure

The data for this project is paradata from the Norwegian LFS. It is based on information from both the data collection stage of the survey together and register variables. Only register variables that are available during the data collection stage have been used in this investigation, so that there are practical applications for this work. For example, in Norway there is a register containing information on employment status which correlates very well with our variables of interest, but this is only available well after the data collection process is completed and is therefore not relevant in this setting.

The investigation is divided in two main areas: timing of the first contact and contact progression, which utilize data in different ways. Both sets of data excluded subjects that were never rung and those that were not in the target population group (for example those found to be under the age of 15, over the age of 74 or those residing outside of Norway). All analyses used a status of “interview” as the success dependent variable with which the first instance of this is the one used (despite a few instances of interviews being divided into several timeslots for example). Both direct and indirect interviews (where a parent or spouse is interviewed on behalf of the individual) are included in the data; however this is an area that should be analysed further.

The Norwegian LFS is run as a panel survey in 8 waves. To investigate optimal timing of the first contact, we looked at those in the first wave only and have analysed the data within each of quarters 2 to 4 in 2013. The quarter refers to that which the reference week is in, rather than the actual interview date. Variables used in the analysis for timing of first contact are presented in table B1.1 in Appendix B.

For investigating the second contact attempt, an additional variable of whether an appointment was scheduled during the first attempt was also included. In addition, the time, day and result of the second attempt were investigated. The dataset for the second attempt was restricted to those that did not have an interview in the first attempt and did not have a non-response status after the first attempt (these people were not rung again). Additional variables for second contact attempt are presented in table B1.2 in Appendix B.

For investigating how response propensities and indicators change by attempt, all panel waves of subjects were used. A previous response variable was included as well, in addition to many of the variables listed previously. Only quarters 3 and 4 in 2013 were investigated in this section as previous response was not available in the second quarter. Response rates, R-indicators and the key variable employment status were calculated through the data collection period. Variables for investigation of response propensities and indicators are in table B1.3 in Appendix B.

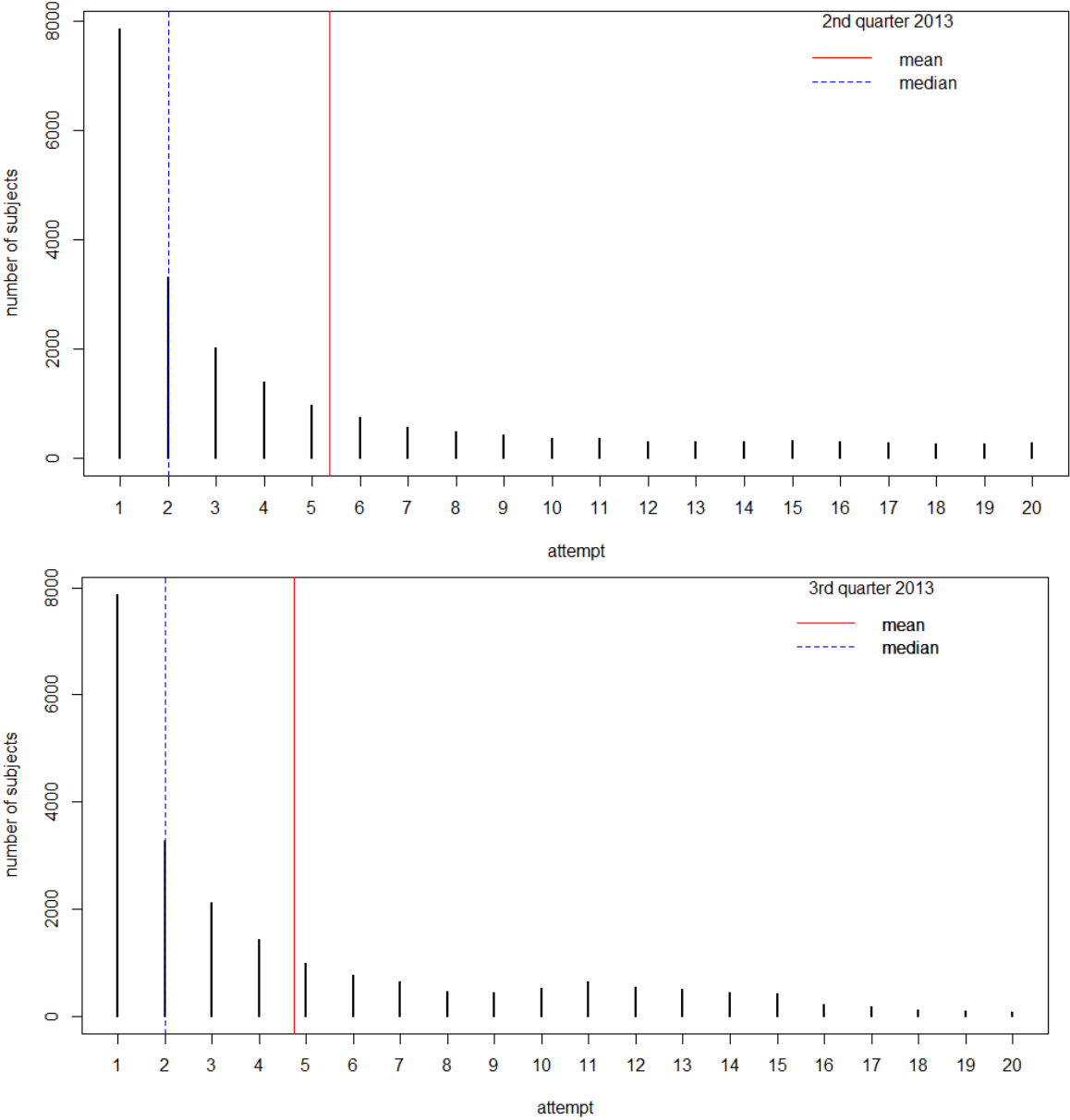
4. Case length

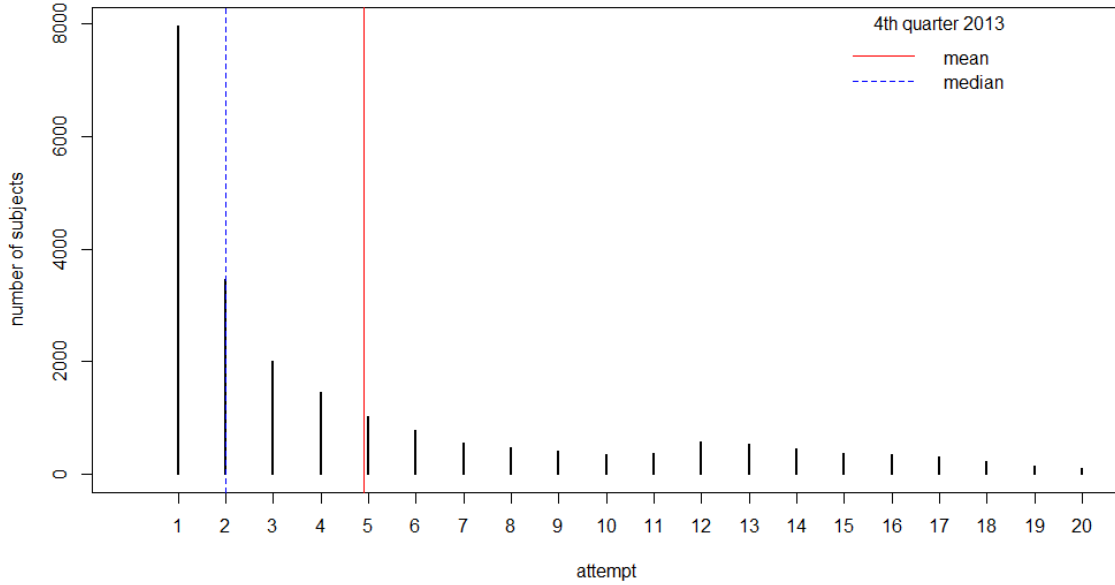
The Norwegian LFS is an important data source for political decisions, and is the only ongoing survey compulsory for individuals. Much time and effort is allocated to ringing and following up subjects in the data collection process. Individuals that are unable to be contacted or interviewed on the first attempt are rung again at a later time or day. There has been a focus on varying the time and day of contact attempts. The number of contact attempts that are made is not fixed, but a case may eventually be

closed without contact being made if the amount of time that has elapsed since the reference week is more than two and a half week (the case for week 13).

We have investigated the first 50 contact attempts. Figure 1 shows the case length distribution up to 20 attempts. In the 2nd quarter there were 931 individuals that were rung more than 20 times. In the 3rd quarter there were 224 and in the 4th quarter 377 individuals were rung more than 20 times. As we see, the total number of attempts required to complete a case can be very high for some cases. The mean number of contact attempts was approximately 5, while the median was 2 for all quarters investigated. This reflects the skew in the data, also shown in the figure.

Figure 1. Number of contact attempts made before a case is finalised by quarter (2-4) in 2013 for LFS





The average number of attempts is higher in the 2nd quarter compared to the 3rd and 4th quarter (table 1). This is to be expected as there are many public holidays in this quarter in Norway when people take holidays and are harder to reach (Easter holiday, May 1st, Ascension Day, the Norwegian Constitution Day on the 17th of May, and Whit Sunday).

Table 1. Average number attempts to finalise a case by quarter in 2013

	2nd Quarter	3rd Quarter	4th Quarter
Mean	5.36	4.74	4.91
Median	2	2	2
Maximum attempts	43	49	39
Total number of subjects	22075	22086	22204

5. Response propensities

The random response model assumes that every person or element k in the population has an (unknown) response probability ρ_k . For the model, there is a set of response indicators R_1, R_2, \dots, R_N where $R_k = 1$ if the corresponding element k responds, and where $R_k = 0$ otherwise. If element k is selected in the sample, a random mechanism is activated that results in response with probability ρ_k and with non-response, probability $1 - \rho_k$.

The unknown probabilities ρ_k can be estimated using the technique with propensity scores. The propensity score $\rho(X)$ is the conditional probability that an individual with the observed characteristics X responds to a survey when invited to participate ($R = 1$):

$$\rho(X) = P(R = 1|X)$$

This assumes that in subpopulations defined by observed characteristics X , all individuals have the same response probability; which implies the "missing at random" (MAR) assumption. Logistic regression or logit models are commonly used to estimate response propensities i.e. the propensity score. Variable choice is always important in these models also when investigating the bias related to a key outcome variable (for example employment status).

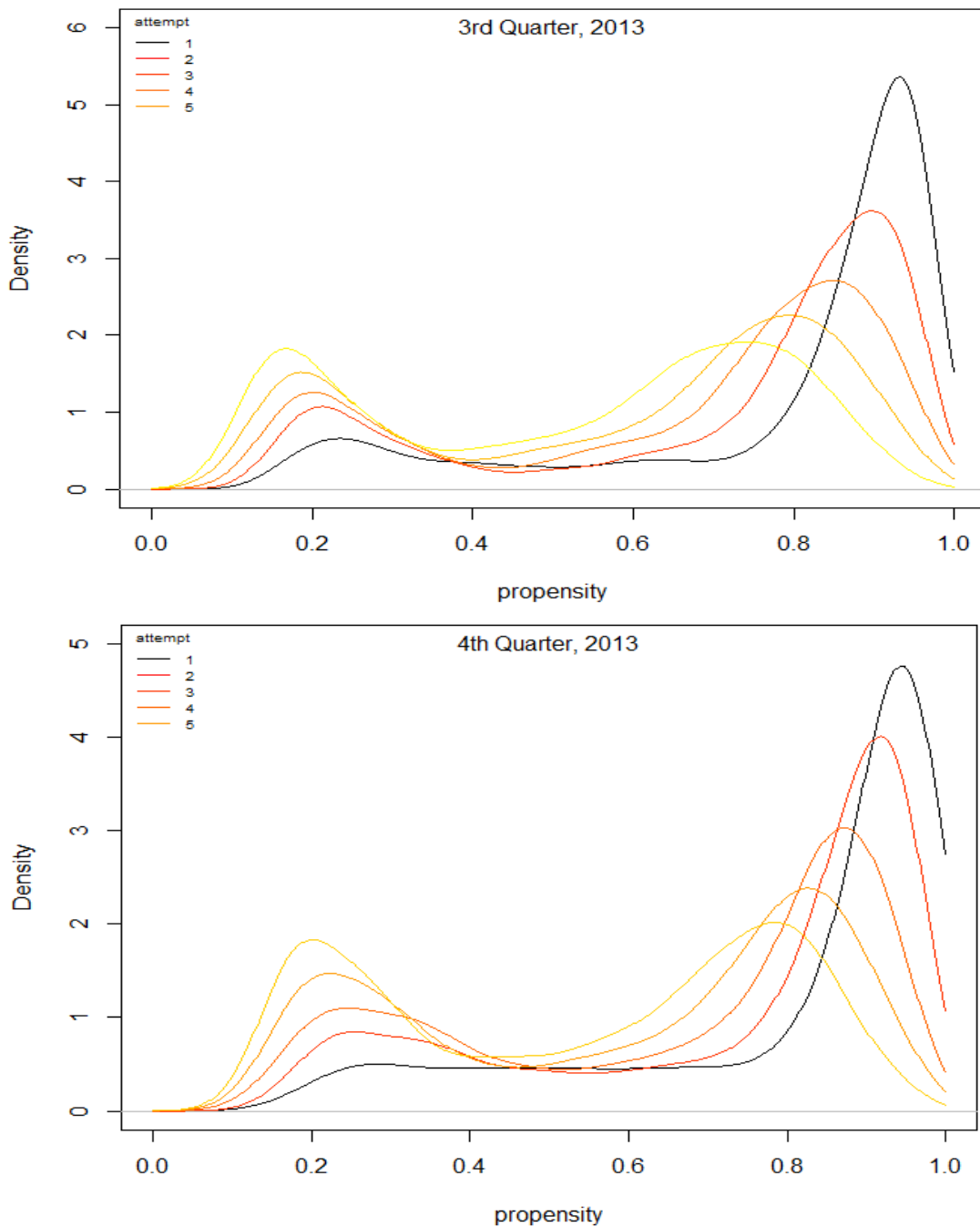
The variables outlined in the previous chapter were used in logit models and reflect both characteristics that are known to effect response propensities and those which we are interested in investigating. They also represent some of the variables which are used for non-response weighting/calibration in the Norwegian LFS. The general formula for a logit model can be written as:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Where p is the probability that an individual responds and $x_{n,\dots,k}$ are binary explanatory variables.

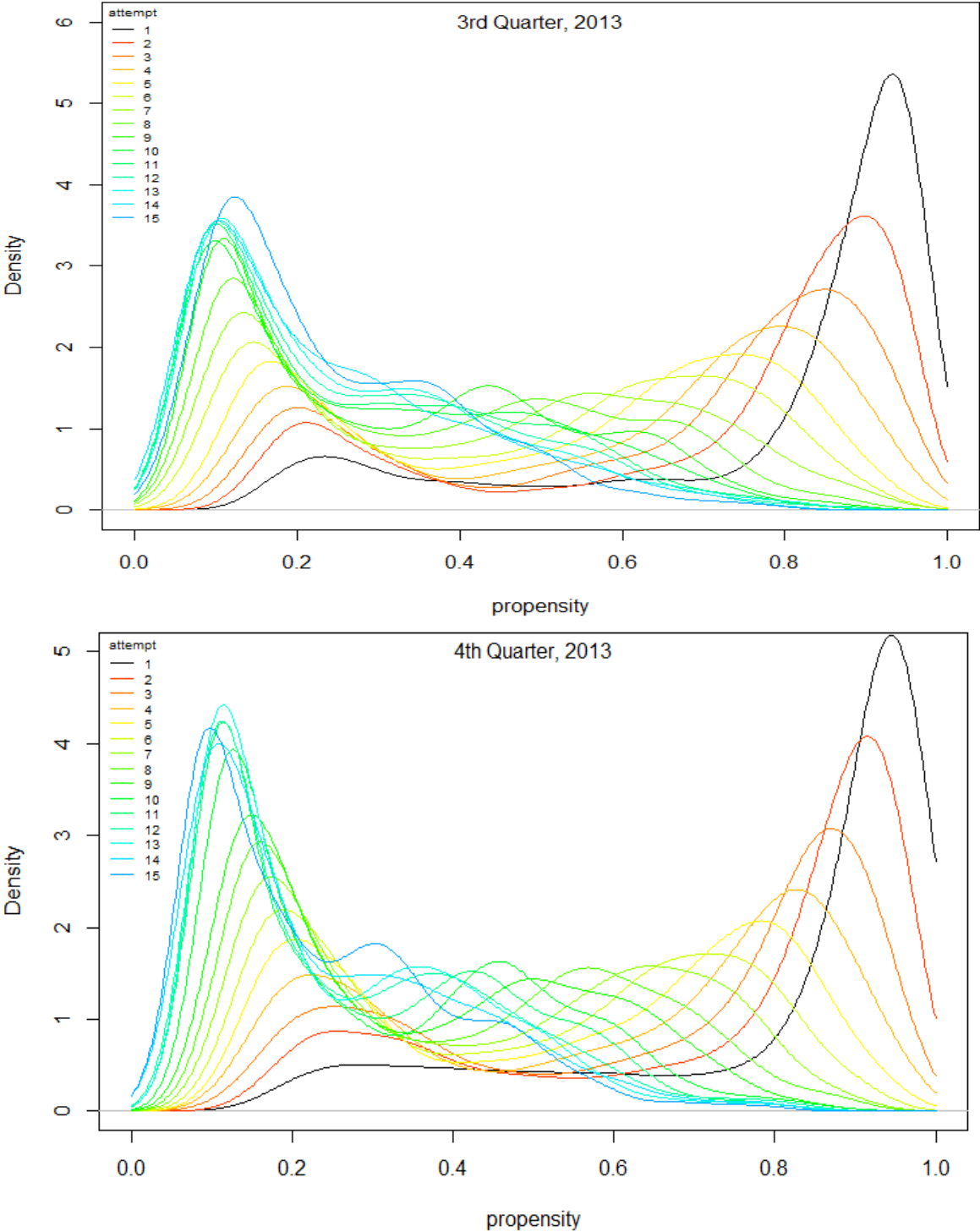
As a starting point, a logit model without interactions was fitted using the variables gender (2 groups), age (7 groups), education level (4 groups), family size (4 groups), last response (6 groups), region (7 groups), time (first contact) and weekday (first contact) for the 3rd and 4th quarter of 2013. The response variable in this case was the final response, i.e. whether or not the subject responded in that quarter. For contact attempts 2-15, whether an appointment had been made in the previous attempt was also included in the model and the corresponding time and weekday were used.

Figure 2. Response propensities for the first 5 contact attempts, 3rd and 4th quarter 2013.



The distribution of the response propensities for the first 5 contact attempts is shown in figure 2 and for the first 15 attempts in figure 3. At the first contact attempt, response propensities are high but with a long tail. The distribution of response propensities becomes bi-modal after the second attempt, indicating two groups of subjects – those that are quite likely to respond and those that are not likely to respond. In general, the distribution of response propensities decreases by contact attempt, i.e. it becomes increasingly harder to make contact and achieve an interview (figure 3).

Figure 3. Response propensities for the first 15 contact attempts, 3rd and 4th quarter 2013.



From figure 3 we see that it might be interesting and valuable to call respondents with response probability in the range 0.4 -0.6.

6. Investigation of timing for first contact

To keep costs as low as possible and for data quality reasons: to avoid recall error, the interviewers at Statistics Norway aim to call the respondents as close as possible to the reference week. Hence it is interesting to investigate the timing of the first and second contact (especially in the first wave).

By additionally investigating the timing of contacts in relation to these problem groups; we can identify ways to maximize their response propensities, which later on could be used in a responsive design framework.

6.1. First contact attempt for first wave of subjects

A full model was fitted to the data including all three available quarters in 2013. The full model included variables (all categorical): gender, age, time, weekday, family size, education, region, priority (a variable that indicates priority in the Blaise system) and quarter. The response variable was whether or not an interview was achieved on the first attempt. All 2-way interactions between age, gender, family size, time, weekday and priority were included. AIC backward selection was used to find the best model. The main effects for region and weekday were excluded from the best fit model. The resulting interactions were included in the model with the best fit:

- *Gender × Age*
- *Gender × Priority*
- *Family Size × Priority*

A post hoc Tukey test (Tukey, 1949) for time was performed to specifically investigate pairwise comparisons of the time levels of this variable (table 2).

Table 2. Number of subjects rung for the first time by time slot. Quarters 2, 3 and 4 2013.

Time slot	before 13:00	13:00-17:00	17:00-19:00	19:00-21:30
Number of subjects rung	402	1578	3491	3352

Time slot	before 13:00	13:00-17:00	17:00-19:00	19:00-21:30
before 13:00				
13:00-17:00	-0.02			
17:00-19:00	-0.00	0.01		
19:00-21:30	0.01	0.03**	0.03*	

*p<0.1

** p<0.05

*** p<0.001

Table 2 gives the difference between the vertical timeslots compared to those listed horizontally. For instance the last timeslot (19:00 – 21:30) is significantly better for contacting people than the second to last timeslot (17:00 – 19:00).

6.2. Best day for contact

When we look at the response rates by weekday, Wednesday first appeared to have the highest response rates for first contact. However the variable weekday falls out of our model when the priority variable is included. This is likely because those rung for the first time later in the week often have a soft appointment on their potential availability made earlier in the week with another family member. Additionally, family members may have already informed them about the survey and act in raising response rates. This is not captured in our data and it means that those rung later in the week for the “first” time should actually be registered as a second attempt. Better use of the *priority* variable (see Appendix B1.1) which indicates if the interview was started automatically or manually may aid in future analyses of this. We have noted some limitations with this variable in the dataset and are unsure how reliable it is to use in this context. For example, we observed that 38 per cent of first time interviews are started automatically on a Friday (table 3), however those dealing with interviewing maintain that everyone is contacted for the first time early in the week. There are perhaps family and other effects at play here too and we suggest further investigation on how to use this variable in analyses.

Table 3. Response rates, subjects rung and per cent interviews started manually.

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Response rate	0.20	0.25	0.42	0.39	0.34	0.32
Number of subjects rung	5129	2281	531	407	357	118
Per cent of interviews started manually	16	36	88	87	62	96

The interviewer system, Blaise, at Statistics Norway is fully loaded with new respondents on Mondays, and most cases are contacted at least once by Tuesday evening every week. To control for the day the system is loaded, we looked at two shorter weeks where Monday is a national holiday (Easter Monday and Whit Monday). In these cases the system will be fully loaded on a Tuesday instead. Unfortunately for the week starting with Easter Monday there were too few observations, but for Whit Monday we found that Friday was significant better than Thursday.

Days later in the week may be better for contacting subjects, but the data collection setup does not allow us to easily study this at the moment for this survey; we lack the necessary paradata for that analysis. We believe this is an area that deserves further attention.

5.3 Second contact attempt for first wave subjects

A full model was fitted including all the variables from the first attempt model, and in addition: time and weekday of the second attempt, and appointment (yes/no). The variable priority was not included as the appointment variable is used here. The response variable in this case is whether or not an interview was achieved in the second contact attempt. Only subjects that did not respond and were not classified as non-respondent in the first attempt were included in the analysis. All 2-way interactions between, gender, age, family size, weekday (first attempt), weekday (second attempt), time (first attempt) and time (second attempt) were included. AIC backward selection was used to select the best model. Region was the only main effect to not be included in the best model. The interactions which were included are:

- *Gender* × *Age*
- *Gender* × *Time (second)*
- *Time (first)* × *Time (second)*
- *Time (second)* × *Weekday (second)*
- *Weekday (first)* × *Weekday (second)*

This indicates a significant interaction between the first and second attempt weekday and timing. This also implies that the “best” weekday and time to contact a respondent for the second attempt is dependent on when the first attempt was, something that could be interpreted as the respondents have preferred weekdays and/or time that day they like to be contacted at. We think this is an area that needs further investigation.

7. Phase capacity and mode switch in the Norwegian LFS

To build a responsive design, it is essential to have a set of indicators to follow in real time during the data collection. The indicators should reflect when during the data collection the value of the survey variable becomes stable. The implication of this is that new respondents may not add any extra information to the estimates, they just drain the budget. This is a phase in the data collection where it is natural to swap mode or to stop the entire collection process. In some cases, for some subgroups of respondents, continuing data collection might just make estimates worse (more skewed). Such state in the process is often referred to as phase capacity where each set of design features brings with it a maximum level of quality given the costs (Groves and Heeringa, 2006).

We have examined when to give up data collection or possibly switch to another mode (e.g. web) using R-indicators and specific groups based on response propensities.

7.1. Representativity and R-indicators

The response to a survey is representative with respect to X when the response propensities $\rho_X(x)$ are constant over x , i.e. they are equal to $\bar{\rho}$ (Bethlehem et. al., 2011). The concept *representative response* is the degree of similarity between the set of respondents and the complete sample with certain socio-demographic or socio-economic characteristics. From this concept two indicators are proposed: the R-indicator and the partial R-indicator. The R-indicator is equal to one minus two times the standard deviation of the response propensities:

$$R(\rho_X) = 1 - 2S(\rho_X)$$

where

$$S^2(\rho_X) = \frac{1}{N} \sum_{k=1}^N (\rho_X(x_k) - \bar{\rho})^2$$

The R-indicator is motivated by the idea, that the better the non-response behaviour can be explained, the further away the indicator is from representativity. But we also know that the better the non-response can be explained, the better it can be adjusted for (in general). R-indicators range from 0 to 1 with 1 being optimal and equal to representative response. The indicators can for instance be used during data collection to draw attention to groups in the population that easily get underrepresented in a survey. If new respondents are to be found, like in our case, we want to get in contact with the ones who never participate, it is best to seek them in those groups that are the most underrepresented.

The partial R-indicators we have used are of a conditional type. They are a measure the contribution of each variable to a lack of representative response given the other variables. The conditional partial R-indicator at the variable level is described further in Bethlehem et. al. (2011). The closer the conditional partial R-indicator is to 0, the less the variable contributes to a lack of representativity.

The conditional partial R-indicators are especially suited for monitoring the data collection, where they can assist the R-indicators in identifying groups that are underrepresented. They can also be used to support decisions in responsive data collection designs, together with maximal bias figures. Bethlehem et. al. (2011) gives further details. When improving the representativity, the objective is always to increase the response rate and decrease the variation in response propensities.

7.2. Our model

The model we fitted to calculate R-indicators included the variables previously used: education, region, family size, last response, gender and age with an interaction term between gender and age.

We have analyzed the stability of employment as a function of call attempts. The estimate for the employment rate in this analysis was not weighted or calibrated. It is therefore important not to compare these results with official figures published by Statistics Norway. Additionally, there maybe be some trend differences between the calibrated and raw estimates, but without further investigation we cannot asses what these differences may be.

The level for the estimate for employment rate appears quite stable after around 5 attempts in both the 3rd and 4th quarter (figure 4). This is expected since the average number of call attempts was just below 5 for these quarters with a median of 2. We can also see the importance of including the last response variable in the model; it adds extra variation in the response propensities, and hence the R-indicator has a lower value of about 0.6, compared to the R-indicator without last response which stabilizes at about 0.8.

Here we have only investigated one interest variable (employment) in the Norwegian LFS, but for a responsive design framework, it is also important to look at other interest variables, for example un-employment.

Figure 4. R-indicators for 3rd and 4th quarter 2013. Estimate for employment rate not weighted or calibrated.

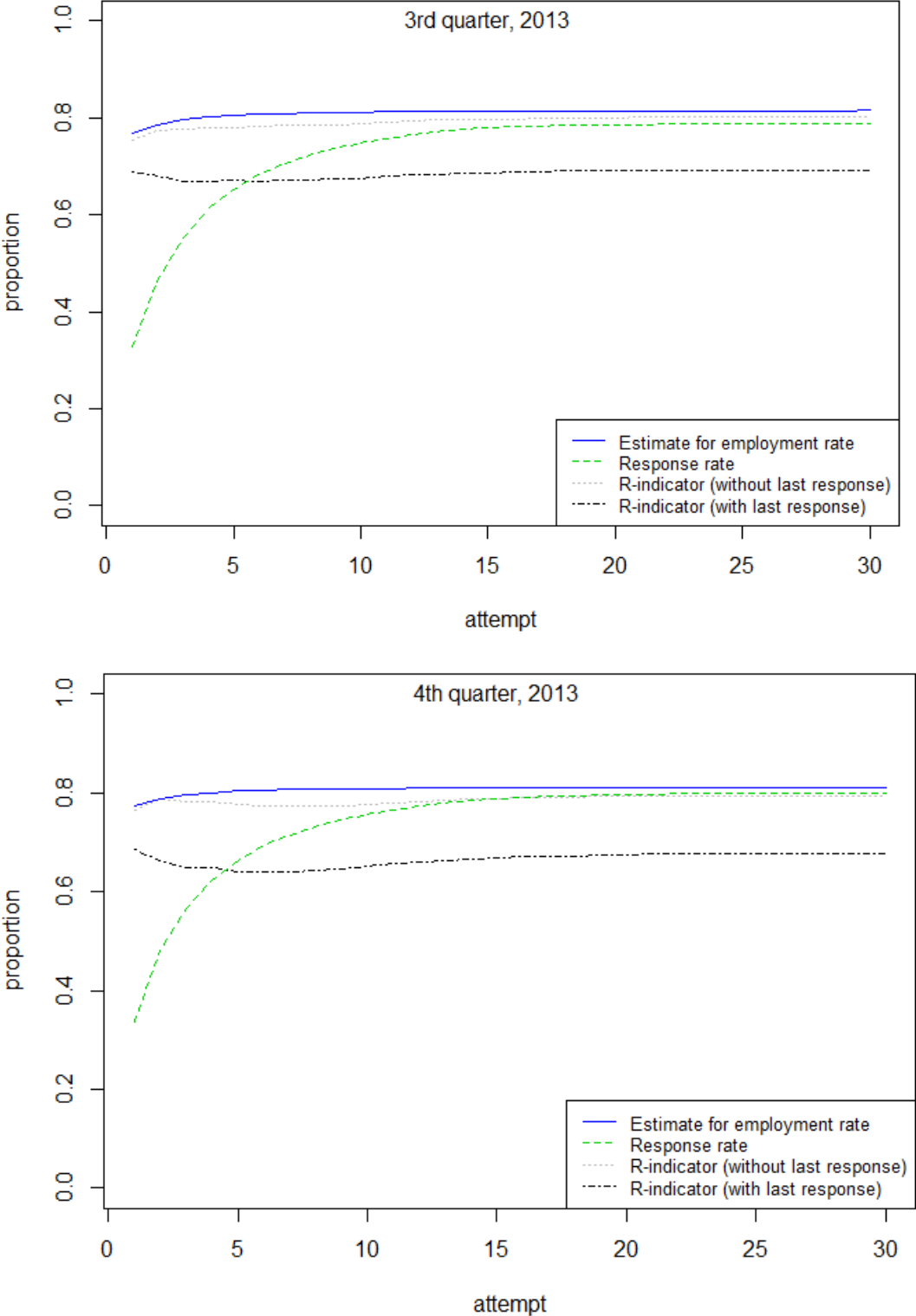
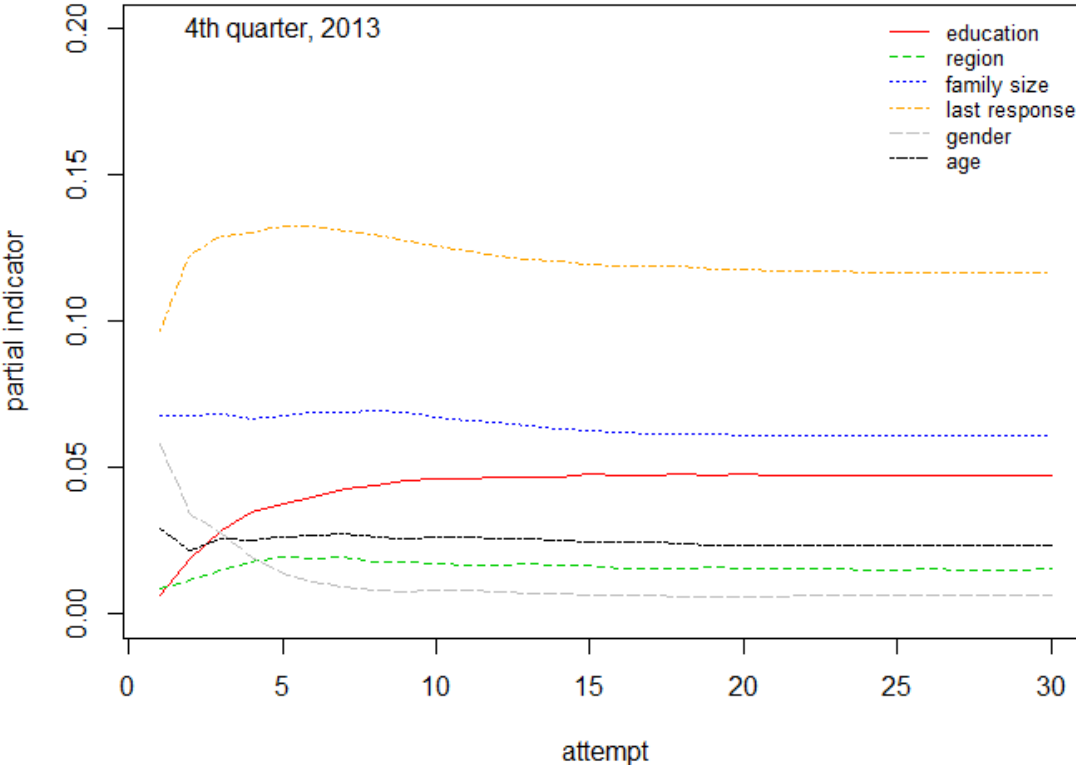
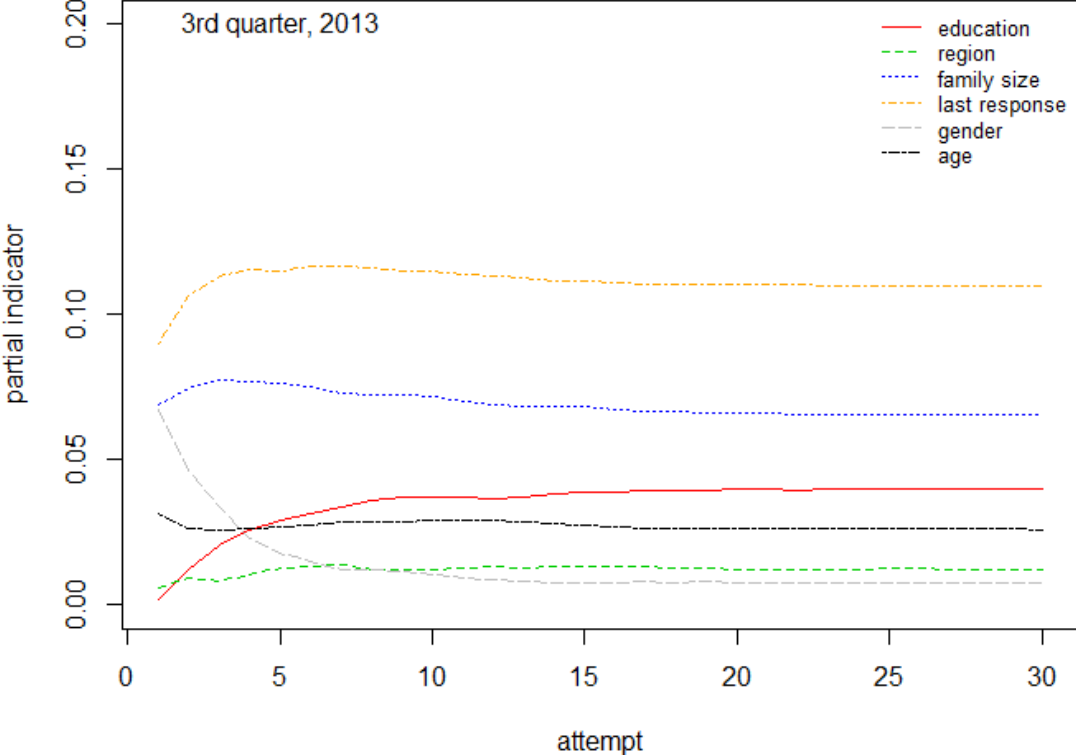


Figure 5 shows small changes in the partial R-indicators for most of the variables in the first few attempts except for the variable last response. For the variables age and gender the representativity improves while education and last response worsens with increasing attempts. For last response we see that we get hold of more and more of those respondents who have participated and answered before in the survey, which really is not optimal.

Figure 5. Partial R-indicators for 3rd and 4th quarter 2013.



The groups that already are doing well makes it even better as data collection continues. The underrepresented groups shine even more by their absence, and the net sample moves further away from representativity. This is a similar result to a previous study by Pedersen and Falnes-Dalheim (2012).

From figure 5 we can also draw the conclusion that all variables stabilizes after 10-15 attempts, indicating there is no point in continuing ringing anymore. This might be the state in the data collection where it is time to do a mode switch or maybe stop collecting data. We think this should be investigated further in the future, preferably with an experiment aiming to measure phase capacity of different modes. The experiment could also be done retrospective on already finished data collections (Lundquist and Särndal, 2013).

It is natural to assume that something characterizes those who always are participating and therefore responded last time. But what is more interesting is the difference between the groups; the respondents and non-respondents. That is to say, what characterizes those who have not participated before and have no “last response” track record. These are the people we really would like to get in contact with, they would contribute and balance our net sample, make it more representative. The characteristics for these persons must be found in the groups with the lowest probabilities to respond: persons aged 25-34 without higher education, belonging to single person household who did not have a booking as an outcome from last contact attempt and was not interviewed in the previous quarter. These findings are echoed in Lagerstrøm et. al. (2015).

Since we have neither managed to establish contact with these people, nor gained cooperation from them, one may wonder if we contact them in the way they want to be contacted. Maybe chasing them on phone isn't the right mode; it just costs money. They could be offered a web survey straight away instead.

8. Concluding remarks and further analysis

8.1. Suggestions for a responsive design

The proposal for a new data collection strategy for the Norwegian LFS is outlined in Gravem et. al. (2014). For the majority of people with simple working hours, the survey would run as an online form with non-respondents followed-up by telephone. A final swap back to the online form is planned for non-respondents unable to be contacted by telephone (Gravem et. al., 2014).

SSB has a web option developed for the LFS as of today, but is not currently used in production. We believe and that developing any missing infrastructure should be prioritised in order to allow a responsive design to be implemented in the very near future.

Our study provides some basic ground work on when an appropriate time to swap to a new mode may be. Two of the ways we may consider the best point to change are in terms of phase capacity and for maximizing survey representativity. One option could be to base the timing of the swap on a responsive design, whereby certain groups swap to the final cheaper mode when the response rates within the group are above a certain level. Similarly, groups could be routed to the more expensive (but more efficient at recruiting subjects to the survey) telephone follow-up sooner, if initial response rates within the group are below a certain level. Simple control charts for this could be developed with basis in our figures in chapter 6, preferably where the response or outcome influence on the estimate is shown. This since we through this basic analysis have shown that some background variables get more skewed after a certain number of contact attempts.

Alternatively, a pre-data collection design could be set-up so that only those with particularly high response propensities would start in the online mode while others would start at the more efficient telephone interview stage. This is backed-up by the idea that respondent groups that have a high response propensity will probably do very well and respond, regardless of mode for contact.

Finally, we have a certain level of nonresponse; people that we never establish contact with even if we have good contact information. Our study has shown that some of these people belong to respondent groups with the lowest response propensities; still we waste a certain amount of money on them on several contact attempts with telephone interviewers putting in a lot of effort. We suggest giving these people (respondent groups) the web-option from start of the data collection instead, since they obviously aren't available to us by phone.

8.2. Further analysis

This short analysis has sparked many more questions and areas for future study. We have currently observed respondents as independent individuals; however, the Norwegian LFS is a household survey. All family members of the selected individual are recruited to the survey. It is assumed that this is an "easier" way to recruit the sample because there is a general idea that when one person has participated, others within the household are more likely to too. It also allows for the possibility of indirect interviews which are common for certain groups (e.g. the youngest) in our survey. This data provides a base for a thorough analysis of this effect and how it work by contact attempt and is an area of future study.

The data collection setup does not allow us to easily study if the days later in the week could be better to start ringing on at the moment. We suggest this is an area that deserves further attention. We recommend a pilot survey where initial contact for day and time vary to allow further analysis of this. This could lead to improvements in the pre-data collection design to optimize initial response rates. In addition, the inclusion of indirect interviews may be investigated in future as we are particularly interested in when we are able to make direct contact with respondents.

When an interviewer works through all possible respondents loaded in the Blaise system, some families are coded as "parked". For instance, a family might be parked due to prolonged illness within the family or if they have been burdened by several surveys in recent years. We would like to investigate these parked families a bit further, identify which group of respondents they belong to (if it is possible to find anything common among them) and examine how many extra phone calls we make to these families as well as how many extra interviews we get on these extra calls on average.

Lundquist and Särndal (2013) have studied balance and representativity indicators in the Swedish Living Conditions Survey. They have investigated how a responsive data collection design could be used to improve the final response set in terms of balance and with great cost reductions. This is done by stopping the data collection within specified groups (for example age, gender categories) when a certain response rate is reached. Our data could also provide a suitable base for future investigations similar to this.

The timing of the mode changes is another area of further investigation and an experimental pilot study should also be conducted. In this context it may also be interesting to investigate when in the actual field period a respondent answers given the response probability for the same respondent.

References

- Bethlehem, J., F. Cobben and B. Schouten (2011). *Handbook of Nonresponse in Household Surveys. Wiley Handbook in Survey Methodology*. New Jersey.
- Biemer, P. P. and L. E. Lyberg (2003). *Introduction to Survey Quality. Wiley Series in Survey Methodology*. New Jersey.
- Gravem, D. F., T. Båshus and B. O. Lagerstrøm (2014). The organisation of mixed-mode data collection. ESSnet Data Collection for Social Surveys using Multiple Modes. Deliverable WPIII.
- Groves, R. and S. Heeringa (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, **169** (3), 439-457.
- Laflamme, F. (2009). Experiences in Assessing, Monitoring and Controlling Survey Productivity and Costs at Statistics Canada. Proceedings of the 57th Session of the International Statistical Institute, South Africa. (August 16-22).
- Lagerstrøm, B.O., M. Lillegård and T. Löfgren (2015). A new data collection strategy based on respondent groups and conditional response rates. A case from the Norwegian Labour Force Survey. Paper presented at the International Workshop for Household Nonresponse, Leuven September 2015.
- Lundquist, P. and C. Särndal (2013). Aspects of Responsive Design with Applications to the Swedish Living Conditions Survey. *Journal of Official Statistics*, **29** (4) 557-582.
- Mohl, C. and F. Laflamme (2007). Research and responsive Design Options for Survey Data Collection at Statistics Canada. In proceedings from Joint Statistical Meeting. Available at: <http://www.amstat.org/sections/srms/Proceedings/y2007/Files/JSM2007-000421.pdf>
- Pedersen, H. E. and E. Falnes-Dalheim (2012). Non-response and representativity in a survey on education completed abroad. Q2012 Conference in Athens in May 2012. Documents 2012/38, Statistics Norway.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Tukey, J. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics*, **5** (2), 99-114
- Wagner, J., B. T. West, N. Kirgis, J. M. Lepkowski, W.G. Axinn and S. K. Ndiaye (2012). Use of Paradata in Responsive Design Framework to Manage a Field Data Collection. *Journal of Official Statistics*, **28** (4), 477-499.

Appendix A

Output for first attempt Tukey test for timeslot

```
$starttidgrp1
      diff      lwr      upr      p adj
2-1 -0.019038258 -0.075102494 0.03702598 0.8190762
3-1 -0.012291304 -0.065144845 0.04056224 0.9328359
4-1  0.014016499 -0.038950083 0.06698308 0.9047441
3-2  0.006746954 -0.023693601 0.03718751 0.9411543
4-2  0.033054756  0.002418352 0.06369116 0.0285265
4-3  0.026307803  0.002040797 0.05057481 0.0274551
```

Output from best model for first contact attempt

Call:

```
glm(formula = result1 ~ utdgr + kvt + starttidgrp1 + kjonnn *
     aldgr3 + kjonnn * innpri + famsize * innpri, family = "binomial",
     data = yywavel)
```

Deviance Residuals:

```
      Min       1Q   Median       3Q      Max
-1.6833  -0.6349  -0.4913  -0.3540   2.4144
```

Coefficients:

```
      Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.15886    0.21975  -5.273 1.34e-07 ***
utdgr2        -0.22371    0.11419  -1.959 0.050092 .
utdgr3         0.14508    0.12237   1.186 0.235794
utdgr4         0.34283    0.12543   2.733 0.006269 **
kvt3          -0.10270    0.06696  -1.534 0.125091
kvt4          -0.15178    0.06752  -2.248 0.024588 *
starttidgrp12  0.40183    0.14416   2.787 0.005314 **
starttidgrp13  0.44258    0.13574   3.261 0.001112 **
starttidgrp14  0.52284    0.13535   3.863 0.000112 ***
kjonnn2       -0.38478    0.21905  -1.757 0.078997 .
aldgr32       -1.14198    0.17487  -6.530 6.56e-11 ***
aldgr33       -1.59779    0.19246  -8.302 < 2e-16 ***
aldgr34       -1.60852    0.21175  -7.596 3.05e-14 ***
aldgr35       -1.28804    0.21221  -6.070 1.28e-09 ***
aldgr36       -1.07852    0.21661  -4.979 6.39e-07 ***
aldgr37       -0.48712    0.21824  -2.232 0.025615 *
innpri1       -0.38167    0.54772  -0.697 0.485907
famsize2      -0.04722    0.09277  -0.509 0.610739
famsize3       0.05986    0.14057   0.426 0.670255
famsize4      -0.16691    0.18209  -0.917 0.359340
kjonnn2:aldgr32 -0.12903    0.24671  -0.523 0.600971
kjonnn2:aldgr33  0.52070    0.24005   2.169 0.030072 *
kjonnn2:aldgr34  0.62391    0.24616   2.535 0.011258 *
kjonnn2:aldgr35  0.46310    0.24470   1.893 0.058422 .
kjonnn2:aldgr36  0.65611    0.25039   2.620 0.008784 **
kjonnn2:aldgr37  0.65451    0.26267   2.492 0.012713 *
kjonnn2:innpri1  0.38785    0.16925   2.292 0.021929 *
innpri1:famsize2  1.69068    0.54878   3.081 0.002065 **
innpri1:famsize3  1.48842    0.56081   2.654 0.007954 **
innpri1:famsize4  1.89938    0.57232   3.319 0.000904 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix B

Table B1.1. Variables investigated in for the timing of the first contact.

Variable	Description	Category levels	Code name
Gender	The gender of the subject	1 = male 2 = female	kjonn
Age group	Age of subject divided into 7 groups.	1 = 15-19years 2 = 20-24years 3 = 25-34years 4 = 35-44years 5 = 45-54 years 6 = 55-64 years 7 = 65 years and over	aldgr3
Family size	The size of the family unit in LFS. This includes only those 15 years and over.	1 = 1 adult 2 = 2 adults 3 = 3 adults 4 = 4 or more adults	famsize
Region	The region in which the subject lives	1 = Oslo and Akershus 2 = Hedmark and Oppland 3 = Sør-Østlandet 4 = Agder and Rogaland 5 = Vestlandet 6 = Trøndelag 7 = Nord-Norge	landsdel
Education	The level of education of the subject	1 = information missing 2 = No schooling/basic schooling 3 = Upper secondary education 4 = Tertiary education	utdgr
Time of first attempt	The time of day which the first contact attempt occurred	1 = before 13:00 (from 09:45 or 11:45) 2 = 13:00-17:00 3 = 17:00-19:00 4 = 19:00-21:30	starttidgr1
Day of first attempt	The day of the week which the first contact attempt occurred	Monday-Saturday	ukedag1
First result	Binary variable for result of first contact attempt	0 = no interview 1 = interview	result1
Priority	Binary variable for whether the interview was started automatically or not	0 = Interview started automatically or through CAPI 1 = Interview started manually	innprioritet

Table B1.2. Additional variables investigated in for the timing of the second contact attempt.

Variable	Description	Category levels	Code name
Appointment	Whether or not an appointment was scheduled at the first attempt	0 = no appointment 1 = appointment	avtalegr1
Time of second attempt	The time of day which the second contact attempt occurred	1 = before 13:00 (fra 09:45 eller 11:45) 2 = 13:00-17:00 3 = 17:00-19:00 4 = 19:00-21:30	starttidgr2
Day of second attempt	The day of the week which the second contact attempt occurred	Monday-Saturday	ukedag2
Second result	Binary variable for result of second contact attempt	0 = no interview 1 = interview	result2

Table B1.3. Additional variable investigated for response propensities and indicators

Variable	Description	Category levels	Code name
Last response	How the subject responded in the previous quarter	0 = First wave (were not part of the survey previously) 1 = Non-respondent in previous round or unmatched 2 = Employed in previous round 3 = Not employed in previous round 4 = Responded in previous round but unsure of employment status 5 = Not matched in previous round	svarsist
Employment status	The employment status given by the respondent in the survey for the current quarter.	1 = Employed 2 = Not employed	sysstat

Statistics Norway

Postal address:
PO Box 8131 Dept
NO-0033 Oslo

Office address:
Akersveien 26, Oslo
Oterveien 23, Kongsvinger

E-mail: ssb@ssb.no
Internet: www.ssb.no
Telephone: + 47 62 88 50 00

ISBN 978-82-537-9247-7 (electronic)



Statistisk sentralbyrå
Statistics Norway