Statistics Finland

Task force on quality of BCS data: Analysis of response rates

October 2013

Statistics Finland

Statistics Finland

Table of contents

1 INTRODUCTION 3

2 THEORETICAL CONCEPTS ABOUT SURVEY RESPONSE RATES AND UNIT NONRESPONSE 3

2.1 HOUSEHOLD SURVEY RESPONSE RATES: DEFINITION AND TRENDS 3
2.2 UNIT NONRESPONSE 6
2.2.1 Noncontact 6
2.2.2 Refusal to cooperate 7
2.3 SURVEY DESIGN FEATURES AFFECTING RESPONSE PROPENSITY 8
2.3.1 Reducing noncontacts 8
2.3.2 Reducing refusals 8

3 AN INTERNATIONAL COMPARISON OF RESPONSE RATES BETWEEN DG ECFIN CONSUMER SURVEYS 11

3.1 RESPONSE RATES 11
3.2 LINKS WITH DATA BIAS AND VOLATILITY 15
3.2.1 Tracking performance 15
3.2.2 Volatility 18

4 CONCLUSIONS 19

5 REFERENCES 21

1 Introduction

In the framework of the Joint Harmonised EU Programme of Business and Consumer Surveys the 28 Member States of the European Union with candidate countries Turkey, Montenegro, the former Yugoslav Republic of Macedonia and Serbia collect monthly information on households' spending and saving intentions and on the possible factors affecting these decisions. The four main topics of the survey are the households' financial situation, the general economic situation, savings and intentions with regard to major purchases. (See DG ECFIN 2013)

The focus of this paper is on the above-mentioned EU consumer surveys. Like in any other household survey one of the most interesting and important questions is the quality of survey data and the quality of survey estimates. The importance of survey data quality was also recognized at the EU Workshop on Business and Consumer Surveys in November 2012, where it was first discussed the possibility to launch a task force looking deeper into these things. The task force was launched in the spring 2013 with several topics to be investigated. The purpose of this paper is to analyse the response rates of consumer surveys and their effect on the data quality. The leading institution of this thematic group is Statistics Finland.

This paper begins with a statistical literature view in Chapter 2 about survey response and nonresponse. The purpose of the statistical theory is to give an overall view about the factors affecting survey participation, i.e. what are the most familiar reasons for (unit) nonresponse and how the response rate could be increased. The focus of Chapter 2 is in two important reasons for nonresponse, namely the problem of not contacting the sample person and the problem of non-cooperation. The results of the empirical analysis of this thematic group are presented in Chapter 3. Conclusions and recommendations are given in the final Chapter 4.

2 Theoretical concepts about survey response rates and unit nonresponse

2.1 Household survey response rates: definition and trends

The statistical literature about response rates, nonresponse and nonresponse bias is extensive, the dominating goal mainly being the maximization of response rates. However, in recent years there has been much effort to recognize the nature of nonresponse and its effect on survey quality. At present, it is well known that nonresponse bias of a sample mean, for example, is dependent on the response rate and on the values of the target variable. Thus, if the values of the target variable are somehow different between respondents and nonrespondents, the estimate basing only on respondents might differ from that based on the total sample. According to a modern view, the stronger the relationship between the target variable and the likelihood of respond, the larger is the nonresponse bias (Groves et al. 2009).

In this paper, we are only studying the effect of *unit nonresponse*, i.e. the situation where the selected sample person does not provide any information at all to the survey. Consequently, the number of participants will be smaller than the original sample size. The impact of unit nonresponse is seen at the response rate of the survey, which can be regarded as an indicator of survey quality. Response rates are frequently

used to compare survey quality between countries and often a specific response rate target is set to achieve a certain quality level.

The only problem is that there seems to be no internationally accepted standard definition for the response rate. There is some effort to develop standards, the recent work of the American Association for Public Opinion Research (AAPOR 2011), for instance. The AAPOR recommendations concern random digit dialling (RDD) telephone surveys, in-person household surveys and mail or internet surveys of specifically named persons. These recommendations are especially useful when the survey involves a single respondent within a household and follows a sampling procedure common in the United States.

Response rate calculation begins by defining the initial sample size I to be

$$I = NC + OC + RF + NA + R,$$

where NC denotes the number of noncontacts, OC the number of non-eligible contacts (overcoverage, e.g. deaths), RF the number of refusers, NA the number of not-able respondents and R the number of respondents. According to Bethlehem et. al (2011, p. 11) the simplest definition of a *response rate* would be the ratio of the total number of respondents divided by the number of eligible contacts in the sample:

$$Response\ rate = \frac{R}{E}. \tag{2.1}$$

However, defining the number of eligible sampling elements is not always so straightforward. For example, if the sample of a household survey is selected from a frame consisting of telephone numbers with RDD, some numbers selected may belong to businesses, may be out of use or do not belong to the target population for some other reason. This requires a screening step to solve out the eligibility of the sampled numbers. With some numbers it is possible that no-one answers even after repeated calls and the survey organisation ends up being unaware whether the number not contacted is eligible or not. This results in a situation where it is not known how many of those noncontacts are eligible.

If it was supposed that all noncontacts were eligible, the total number of eligible contacts E would be the initial sample size I minus the number of non-eligible sample elements (deaths, emigrants, etc.). The response rate presented in equation 2.1 would simply be

$$Response\ rate = \frac{R}{NC + RF + NA + R}. \tag{2.2}$$

However, the assumption about the all noncontacts being eligible is not always realistic. This is especially the case with random digit dialling. Another possibility is to suppose that the proportion of eligible sample elements is the same among noncontacts and contacts. The response rate would then be equal to

$$Response\ rate = \frac{R}{NC\,\frac{RF + NA + R}{OC + RF + NA + R} + RF + NA + R}, \tag{2.3}$$

where OC denotes the number of non-eligible contacts, i.e. elements not belonging to the sample population. (Bethlehem et. al 2011)

The consumer surveys are usually done with a straightforward method with only one person per household being interviewed. For these kinds of surveys, the response

rates given in equations 2.2 and 2.3 are possible to be used. However, if there were more than one interview per household, the situation would be more complex. Another situation requiring consideration is the computation of the response rate of a self-administered survey. If the self-administered survey bases on a proper sample selection scheme, the response rate can be calculated as the ratio of the respondents and selected sample elements. However, if the self-administered survey is also a self-selection survey, where no sample is selected and the survey questionnaire is simply put on the Internet, the response rate cannot be calculated at all.

According to several sources, there is some evidence from declining response rates over time in many countries. Curtin et al. (2005) showed that the response rate of the Survey of Consumer Attitudes (SCA) conducted by the University of Michigan declined from 72 percent to 60 percent between 1979 and 1996. The response rate of the SCA was only 48 percent in 2003, describing further deterioration since 1996. Between 1996 and 2003 the growing rate of overall nonresponse has mainly been a consequence of an increase in the amount of refusals. A cross-country comparison of de Leeuw and de Heer (2002) showed that response rates had been declining and that there were differences in the nonresponse trends between countries. They also found that the rate at which the refusals are increasing differs between countries while the noncontact rate is increasing at the same phase in different countries. Thus, the increasing trend in the nonresponse rate was caused by the increasing amount of refusals.

In the framework of the Task Force Statistics Finland collected information on the response rates of the consumer surveys between Business and Consumer Survey (BCS) countries[1]. There were some countries reporting declining response rates over the years but also countries reporting stable or even increasing response rates. Countries reporting decreasing response rates were Austria, Bulgaria, Poland, Slovenia and Finland. For detailed information about response rates in Finland see Figure 2.1. France also reported that the response rate has recently slightly decreased. In the Netherlands, the response rate had been declining over time but has recently stabilised. Countries reporting stable response rates were Denmark, Greece, Latvia and Lithuania. Cyprus reported gradually increasing response rates between 2008 and 2013 and Portugal an increase beginning from 2012.

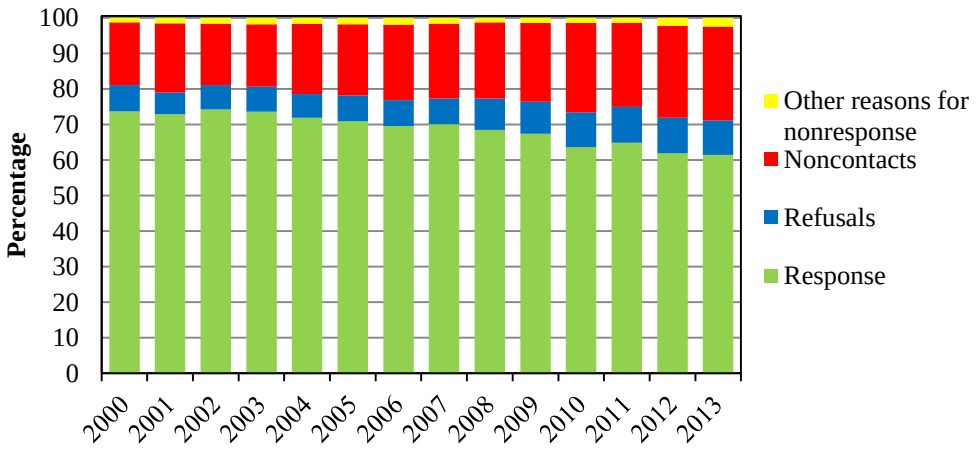


Figure 2.1. Response and nonresponse percentages for the Finnish Consumer Survey

[1] The response rate of a quota sample is usually 100 percent, i.e. the response rate of a quota sample does not usually take nonresponse into account. Hence, developments in the response rates were asked only if the BCS country had a probability based sampling method (not quota sampling method).

2.2 Unit nonresponse

Statistical theory separates between the situation where the selected sample person does not provide any information at all to the survey and the situation where selected sample person answers only some survey questions but not all. The former is defined as unit nonresponse and the latter as item nonresponse. The response rate is affected by the unit nonresponse only and hence this paper focuses only on unit nonresponse.

Usually unit nonresponse is discussed in three categories having distinctive causes:

1. The selected sample element is *not contacted* due to failure to deliver the survey request
2. The selected sample element *refuses to cooperate*, i.e. refuses to participate in the survey
3. The selected sample element is *not-able to participate* in the survey (because of language problems, illness, etc.)

The first and the second reason are discussed further in this paper whereas the "not-able to participate" group is only mentioned here. This is because the not-able category usually does not have very large effect on the overall nonresponse. In addition, nonresponse in the not-able category cannot usually be affected by the means of survey design features, which are important from the viewpoint of this Task Force.

2.2.1 Noncontact

The problem of noncontact arises in interviewer-assisted surveys if the sample elements are not present or are inaccessible to the interviewer. A model presented by Groves and Couper (1998) notes that contactability is affected by the presence of physical impediments, at-home patterns of the household and the timing and number of contact attempts to the sample element. Naturally, these causes differ between survey modes: at-home patterns affect greatly to the likelihood of contact in face-to-face surveys whereas they might not affect so much to the contactability of telephone surveys made mainly into mobile phones, for instance. In face-to-face surveys, physical impediments preventing the visit include locked gates, a locked central entrance to an apartment building, or intercoms. In telephone surveys impediments include answering machines, caller identification features or call blocking services.

The likelihood of contact is affected by household characteristics and survey design features. The household characteristics and socio-demographic factors contributing to survey contactability have been studied by e.g. Groves and Cooper (1998). They found empirical support that locked apartment buildings and high-security housing developments complicated and slowed down contact by requiring more calls to attain the first contact. However, they found no proof that mentioned physical impediments had affected greatly to the overall probability of contact. They also found that older people and households with young children were more easily contacted and that evening calls were more productive than daytime calls.

Stoop et al. (2010) mention the possible influence of the social environment to the socio-demographic characteristics of sample elements and to the presence of physical impediments. It is probable that the general willingness to open the door to interviewers during the evening or the acceptability of unannounced home visits, for instance, differ between countries because of different social environments. Social

environments with higher crime rates may increase the need for alarm systems and security devices, which act as physical impediments to the interviewer. The same reasoning applies to telephone surveys in which it is possible that answering to a call coming from an unknown number of the interviewer differs between countries.

2.2.2 Refusal to cooperate

Refusing to participate is a conscious decision influenced by sociological and psychological reasons. Groves and Couper (1998) describe the request for participate in a survey as a social activity and underline the importance of the interaction between the interviewer and the sample person. They represent a conceptual model for survey cooperation with four different levels influencing on the survey participation:

1. The social environment
2. The person level (householder)
3. The interviewer level
4. The survey design level

As an example of the effect of the social environment Groves and Couper (1998) found that densely populated urban areas tend to have lower cooperation rates. They also note that households with children or with more than one adult tend to cooperate better than single person households do. The effect of interviewer level was seen in higher cooperation rates of more experienced interviewers.

There are also other findings about household characteristics affecting the likelihood to cooperate. Stoop et al. (2010) found in their literature view that the survey cooperation was frequently linked with sex, age, socio-economic status, urbanicity and belonging to an ethnic minority group. They also found suggestions that people with a lower education level or from lower socio-economic backgrounds have higher refusal rates. However, these findings only apply for the analysed surveys. It is always possible that identical neighbourhood characteristics affect differently in different countries. For example, living in densely populated urban areas may reflect social isolation in some countries or cities, but not in others. The role of social environment and survey culture is important also.

One possible theory about influences on survey cooperation is represented by Groves et al. (2000). Their theory bases on assumption that individual persons place different importance on survey design features (e.g., survey sponsorship, topic of the survey or length of the interview). The importance of a design feature is determined by socioeconomic characteristics of the sample person. When the interviewer makes his/her recruit attempt some of these features are more salient to the sample person than others. The final decision to cooperate with the survey request depends on what is made salient to the sample person by the interviewer and how much the sample person negatively or positively values the attribute. This theory is called the "leverage-saliency" theory (Groves et al. 2000).

2.3 Survey design features affecting response propensity

2.3.1 Reducing noncontacts

The household and person characteristics mentioned in Section 2.2.1 have an important role in the survey cooperation but they cannot be directly affected by the survey or fieldwork organization. The only way for survey organization to reduce the number of noncontacts is through different survey design features. This section represents some survey factors known to have an effect on the noncontact rate of an interviewer assisted survey.

Number and timing of contact attempts

Increasing the number of contact attempts obviously increases the likelihood of finding someone at home or the likelihood that someone answers the interviewer's call. Groves et al. (2009) note that the most productive time to contact household in the United States is on Sunday to Thursday evenings and during the day on weekends. Generally, the call scheduling is regulated by the fieldwork organization with specific rules for the calling patterns for households. The rules can be executed e.g. with the help of computer software which distributes the calls to the interviewers in computer-assisted interviewing system. For example, there could be a minimum number of contacts attempts to be made before a sample element can be regarded as noncontact.

Length of the data collection period

The likelihood of contact increases also if the data collection period is extended. For example, fieldwork period limited to only one day and to the daytime of that day results into noncontact with households not having someone at home in that particular time point (e.g. working single person households). The number of possible contact attempts increases obviously with the length of the data collection period.

Workload of the interviewers

The workload of the interviewer is defined by the number of cases they have been assigned to complete during the fieldwork period. It depends e.g. on the length of the data collection period, survey mode, the length of the interview and the nature of the study population. If the workload of the interviewer is too high (the number of cases assigned per interviewer is too high), the interviewer is not able to give enough attention to hard-to-contact sample persons or decide so freely the number and timing of contact attempts. These problems may be overcome by lowering the workload of the interviewers by e.g. employing new interviewers, lengthening the data collection period or changing the survey contact mode.

2.3.2 Reducing refusals

This section gives on overview about some concrete methods to be known to increase the cooperation rate of a survey.

Advance and persuasion letters

A letter informing sampled persons about the upcoming survey and its topic is called an advance letter. It may contain information about the fieldwork organization, survey sponsorship, sampling procedure (why is the sampled person selected) or interview duration, for instance. The advance letter may help the work of the interviewer by

making the sampled person aware about the forthcoming approach of the interviewer: a call/visit of an interviewer is not confused to a sales attempt. The letter may also reduce the need for interviewer explaining who they are and what the survey is about. Even though the advance letter would not be read, a friendly, businesslike letter can create an image about a contact attempt of a serious enterprise or organisation.

Advance letters presenting information about the survey have been proven effective in mail, in face-to-face and in telephone surveys. The recent meta-analysis of de Leeuw et al. (2007) reported that sending an advance letter improves both the response rate and the cooperation rate of telephone surveys. They also found that the positive effect of advance letters exists with RDD- and list-based probability samples. However, there were clear indications that the effect is much larger in list-based samples. Their analysis was based on 29 studies mainly conducted in the United States. Of course, there are also contrary findings about the effect of advance letters. For example, Singer et al. (2000) found no statistically significant difference between the letter and no-letter conditions in their analysis of the Survey of Consumer Attitudes.

In addition to an advance letter, fieldwork organisation can send a persuasion letter to sample persons refused to be interviewed. If the reason of nonresponse is known (e.g. concern about privacy), the persuasion letter can be tailored so that it especially takes the reason for nonresponse into account.

Respondent incentives

An incentive is a small gift or a small amount of money given to sampled persons in order to gain their participation in a survey. The incentive can be given either before the survey participation or after it. If the incentive is prepaid, it is unconditional. A promised incentive is instead conditional on survey participation.

There are some studies about using incentives in surveys, however, many of those have been conducted in countries where providing incentives is less unusual (e.g. the United States, Canada and the United Kingdom). Singer et al. (2002) found evidence that prepaid incentives enclosed with advance letters increased the response rate of the Survey of Consumer Attitudes. They also found that promised incentives had no reliable increasing effect on response rates in RDD telephone surveys. More extensive meta-analysis of Singer (2002) found that in interviewer-mediated surveys incentives lead to higher response rates independent of the survey mode. She also found that the relationship between higher response rate and the effect of incentives is not linear: the effect of incentives decreases with the extent of the incentives making e.g. a very large amount of cash as too high incentive compared to the response burden. However, Singer (2002) reminds that the foregoing conclusions are in need of further research.

Survey mode

Survey data collection has traditionally been conducted with a single mode data collection method. Usually interviewer-assisted data collection methods produce better response rates than self-administered (mail) surveys. For example, Hox and de Leeuw (1994) found in their meta-analysis that face-to-face surveys had the highest response rates, the telephone surveys the next highest and the mail surveys lowest.

The high response rates of face-to-face surveys are related to the interaction between the interviewer and the respondent. Face-to-face surveys produce generally very high

quality data but they may be expensive to carry out. Furthermore, in face-to-face interview the behaviour of the interviewer may lead to interviewer effects and just the presence of an interviewer may increase the nonresponse rate of very sensitive questions (social desirability bias).

Some weaknesses of the face-to-face interview are improved when the survey mode is switched to telephone interview. Firstly, a telephone interview is not experienced as such a personal way to communicate than face-to-face interview. Hence, the bias accounting from social desirability and the variance created by interviewer effect is smaller in telephone surveys than in face-to-face surveys. In addition, telephone interviewing is less expensive than face-to-face interviewing because of better cost efficiency (e.g. no travel costs and shorter interviews).

Of course, telephone interviewing has also some disadvantages compared with face-to-face surveys; it does not enable any visual assistance in the answering situation, it requires the questions to be simple and leaves persons without telephone number outside the survey. Another difficulty lies in the unlisted telephone numbers if the survey is conducted by linking numbers to the sampling frame. This problem can be overcome by random digit dialling (RDD), but then the possible register information from the sample frame is lost.

Self-administered surveys by mail or by Web solve the problem of interviewer effect and social desirability bias. A mail questionnaire is easy to answer whenever the respondent has time for it and it may result in answers that are more thoughtful. In addition, mail questionnaire is the most inexpensive data collection method compared with interviewer-assisted surveys. However, the questionnaire may result in higher amount of incomplete forms (the risk of item nonresponse grows), it is heavily dependent on the visual design and usually the response rate is low compared with face-to-face or telephone surveys. Furthermore, the data collection period cannot be very short and it takes more time to record the data into an electronic database.

Web surveys do not differ greatly from mail surveys. They involve more questionnaire design choices than mail forms and require less time to record the collected data. Nevertheless, they also require an internet access and may create some concern about security issues among respondents.

By mixing and switching modes it may be possible to optimize resources to reduce noncooperation. It is possible to begin the data collection with some cheaper survey mode (e.g., mail or Web questionnaire) and then use more expensive mode (e.g., face-to-face or telephone) for collecting the information from the nonrespondents of the first mode. There are some findings suggesting that the use of a second sequential mode may improve survey response rates (e.g. de Leeuw, 2005 and Dillman et al. 2009) but also contradictory findings (see Junes and Simpanen, 2013 and about four mode survey Beukenhorst and Wetzels, 2009, cited in Bethlehem et al. 2011, p. 113). In addition, Dillman et al. (2009) found that although the response rate was improved with the adaptation of a second mode, the nonresponse error accounting from respondent demographics was not reduced.

The role of the interviewer

It is obvious that the interaction between the interviewer and the sample person is an essential element of the survey participation. In addition to sociodemographic factors of the sample unit and the survey design features, attributes of the interviewer may

affect greatly to the decision to participate to the requested survey. For example, some sample persons may be more interested to cooperate if they believe that participating is a scarce, limited opportunity. If the interviewer notices this in the interaction situation, he/she can try to underline the scarcity point of view by saying e.g. "only one in every 100,000 persons is contacted".

The interviewer's adaptation of his/her approach to the sample person is called tailoring. Tailoring is not limited to just words, it is about all features of communication: it can be related to intonation, to pacing and to volume of speech or to physical movements (e.g. dressing up for an interview in a rich neighbourhood). In addition to tailoring, it is important that the interviewer is able to maintain the conversation with potential respondents so that the probability of obtaining participation maximises.

In interviewer assisted surveys (face-to-face or telephone) the interviewer can play an important role in increasing cooperation rates. A well-known fact is that experienced interviewers tend to achieve higher cooperation rates than their less experienced colleagues (see e.g. Groves and Cooper, 1998). According to Bethlehem et al. (2011) interviewers considering the survey as uninteresting or containing too sensitive questions obtain higher nonresponse rates.

The fieldwork organisation is usually responsible for the training of the interviewers. Bethlehem et al. (2011) advise that the training should be focused on minimization of refusals by tailoring and maintaining interaction. When requesting participation in a survey, a good interviewer is able to notice some phrases and cues that help he/she to self apply appropriate phrases to obtain participation. Of course, sometimes there are situations in which it is important not to push too hard toward a participation decision but instead retreat and contact the sample person again later.

After training the interviewers the fieldwork organization should also monitor the achievements of the interviewers. This is important for recognizing further training requirements and for studying the interaction between the interviewer and the respondents. For example, Statistics Finland's Telephone Interview Center (the CATI Center) regularly records some interviews and monitors the interviewing techniques used and the speed, volume and the intonation of speech. The CATI Center also collects and studies statistics e.g. about the amount of perceived interviews per interviewer and the proportion of refusals per interviewer.

3 An international comparison of response rates between DG ECFIN Consumer Surveys

3.1 Response rates

As it was mentioned in the introductory Chapter 1 of this paper, the 28 Member States of the European Union with candidate countries Turkey, Montenegro, the former Yugoslav Republic of Macedonia and Serbia collect monthly consumer survey data in the framework of the Joint Harmonised EU Programme for Business and Consumer Surveys (BCS programme). This section is about the differences in the response rates between consumer survey countries. A reader is advised to keep in mind the limitations regarding comparisons between different countries. Even though we can compare some survey design features and their effects on response rates, it is probable

that some part of the variation is explained by cultural and socioeconomic reasons. In addition, the presented discussion gives just an overview about the response rate variation. A more precise and thorough analysis would require more detailed information on the response rates, including information on the different forms of nonresponse (e.g. noncontacts and refusals).

A file of metadata received from DG ECFIN and an additional questionnaire sent by Statistics Finland were starting points to the empirical analysis of response rates. A graphical presentation of response rates is presented in Figure 3.1. The smallest response rates presented are under 10 percent whereas the largest response percentages equal to 100 percent. The mean response percentage is about 71 percent almost equalling the median response percentage of 70.

According to the additional questionnaire and the metadata received from DG ECFIN, the one hundred percent response percentages of Estonia, Germany, Hungary, Italy, Sweden and Slovakia are results of quota sampling method. In these countries, respondents are interviewed in such a way that the distribution of selected auxiliary variables matches the population distribution of these variables. Compared with stratified probability sampling, quota sampling does not require any sampling frame from which in advance regulated amount of sample elements should be selected. The interviewers' job is to interview persons until the quotas are filled. The response rates usually do not consider the interview process, i.e. the rates describe the degree of fullness of the quotas not including the nonrespondents. Hence, the response percentages of quota sampling countries should not be compared with probability sampling countries.

In addition to above mentioned quota sampling countries, Luxembourg, the United Kingdom and Ireland also reported using quota sampling method[2]. However, in these countries the response rate is below 100 percent (in the United Kingdom and Ireland the response rate is around 5 percent). Statistics Finland found no explanation to the below 100 percent response rates of these three quota sampling countries. Thus, they are also excluded from the further analyses.

When the quota sampling countries with Belgium, Czech and Malta[3] are removed from the average computations, the mean response rate in probability sampling countries equals to 64 percent. The median response percentage is 65. However, these computations do not consider the different methods for calculating the response rate. Most of the probability sampling BCS countries seem to calculate the response rate as a ratio of the number of respondents to the total sample selected. This type of calculation method equals to equation 2.1 presented on page 4, with eligible persons being all the sampled elements. Another popular method for calculating the response rate among BCS countries is to exclude noncontacts from the calculation of response rates by applying equation 2.2 with the denominator consisting only of refusals (RF) and respondents (R).

[2] In a questionnaire conducted by Statistics Finland also Montenegro, Macedonia, Romania and Croatia reported applying a quota sampling method. Unfortunately, this is probably a result of failed questionnaire design. According to the metadata received from DG ECFIN the sampling method of these countries seems to be stratified probability sampling.

[3] The response rate of Belgium is always 100% if the targeted amount of interviews is achieved. The response rate of Czech is 100% percent although they apply probability sampling. Malta reported not calculating response rate at all.

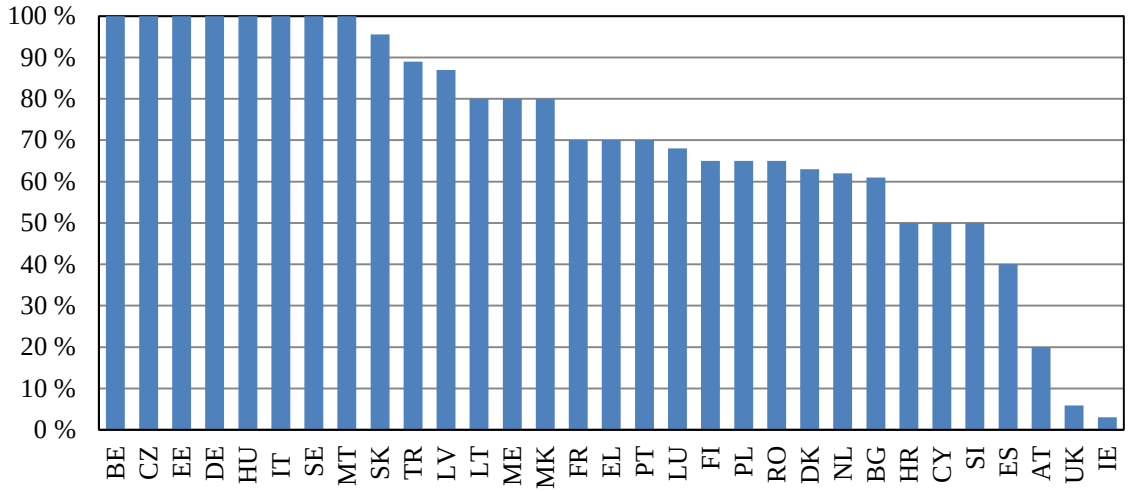


Figure 3.1. Response rates in consumer surveys

BCS countries reporting their response rates as a ratio of respondents to the total sample selected or to the total amount of contacts attempts made are presented in Table 3.1. The way to calculate the response rate is seen in the "Denominator" column, from which the reader is able to verify the denominator used in the response rate calculation. It is here supposed, that the "total attempts made" denominator used by Cyprus and Spain includes also noncontacted sample persons. Hence, the denominator of Cyprus and Spain probably almost equals the original sample size.

The two mixed-mode surveys and the only face-to-face (CAPI = Computer Assisted Personal Interview) survey in Table 3.1 seem to produce higher response rates than most of the telephone surveys (CATI = Computer Assisted Telephone Interview), although the response rates in France and Greece are also very high. Both mixed-mode surveys include a face-to-face data collection period. Thus, the theoretical hypothesis about telephone surveys producing lower response rates than face-to-face surveys seems to be reasonable.

The effect of advance letter is mixed. Spain and Cyprus with response rates of 40 and 50 percent do not send a letter in advance, whereas the very high response rate countries Turkey and Lithuania send an advance letter. However, three countries with a response rate of 70 percent have a different practice concerning advance letters: France and Portugal use advance letters but Greece does not. In addition, the response rate of advance letter nonuser Denmark almost equals to the response rates of advance letter users Finland and the Netherlands. Another possible letter to be sent is a reminding letter. However, this is clearly not a popular method among consumer surveys: only two countries use reminding letters (Portugal sends the letter only to e-mail answerers).

The remaining columns in Table 3.1 are related to the work of the interviewers. The number of contact attempts is limited in some countries with certain amount of attempts to be made before the sample element can be regarded as noncontacted. For example, in France and in Greece sample persons are tried to contact five times before marking the person noncontacted. In Finland and in the Netherlands the number of contact attempts is very high by default. However, in practise it is not probable that a sample person is tried to contact 20 times in Finland. The high number of contact attempts ensures that all noncontacted persons are tried to contact to the end of the fieldwork period. Three countries reported having no limitations regarding

the number of contact attempts. For example, Denmark reported the number of contact attempts varying from 10 to 30 per sample person.

Regardless of various practices in the number of contact attempt regulations, this variable seems to have no direct linkage to the response rates of consumer surveys. For example, the 70 percent response rates in France and in Greece are attained by limiting the amount of contact attempts to five. In Lithuania, however, the sample person is tried to contact only three times resulting into higher response rate than in France. In addition, response rate in countries having no limitations in the number of contact attempts varies from 50 to 70 percent.

Another variable relating to the work of the interviewers is the dependency between the interviewers' salary and the number of completed interviews. From the group of the six highest response rate countries, five countries reported paying the interviewers based on the number of completed interviews. All the other countries presented in Table 3.1 pay the interviewers' salary independent of the number of completed interviews.

Table 3.1. BCS countries reporting response rates as a ratio of respondents/total sample

	Mode	Advance letter	Reminder	Number of contact attempts	Wage dependency	Denominator	Response rate
Turkey	CAPI	Yes	No	At least 3 visits	No	Total sample	**89 %**
Lithuania	Mixed-mode	Yes	No	Limited to 3	Yes	Total sample	**80 %**
France	Telephone	Yes	No	Limited to 5	Yes	Total sample	**70 %**
Portugal	Mixed-mode	Yes	Yes	No limitations	Yes	Total sample	**70 %**
Greece	CATI	No	No	Limited to 5	Yes	Eligible contacts	**70 %**
Finland	CATI	Yes	No	20 by default	No	Sample size	**65 %**
Romania	Face-to-Face	No	No	Limited to 2	Yes	Total sample?	**65 %**
Denmark	Telephone	No	Yes	No limitations	No	Sample size	**63 %**
the Netherlands	CATI	Yes	No	At least 12 attempts	No	Total sample	**62 %**
Slovenia	CATI	Yes	No	No limitations	No	Total sample	**50 %**
Cyprus	CATI	No	No	No limitations	No	Total attempts made	**50 %**
Spain	CATI	No	No	Limited to 3	No	Total attempts made	**40 %**

BCS countries calculating their response rates as a ratio of respondents to contacted households[4] are presented in Table 3.2. In these countries the response rate describes the rate of success in interview persuasion, i.e. the rate at which sample persons agreed to participate in the survey if they were contacted. Presumably, this kind of method to calculate response rate results in higher response percentages than in Table 3.1.

In Table 3.2 face-to-face surveys produce clearly higher response rates than the only single mode telephone survey (CATI). Only one country reported informing the selected sample elements in advance with a letter. Reminding letters are not sent by any country, and the salary of the interviewers is at least partly dependent on the number of completed interviews in all countries. The number of contact attempts was asked also from the countries presented in Table 3.2. However, the number of contact attempts is more likely to have on effect on the contactability of the sample element than on the co-operation likelihood. Because the response rate describes only the co-

[4] In Austria, the number of total usable addresses excludes inexistent and non-reachable addresses as well as other dropouts.

operation likelihood on the condition that the sample element was contacted, "number of contact attempts" column is excluded from Table 3.2.

Table 3.2. BCS countries reporting response rates different way

	Mode	Advance letter	Reminder	Wage dependency	Denominator	Response rate
Latvia	Face-to-Face	No	No	Yes	Completed + Refusals	**87 %**
Montenegro	CATI+email	No	No	Yes	Contacted households	**80 %**
Poland	Face-to-Face	Yes	No	Yes	Contacted households	**65 %**
Bulgaria	Face-to-Face	No	No	Yes	Total contacts made	**61 %**
Croatia	CAPI	No	No	Yes	Contacted households?	**50 %**
Austria	CATI	No	No	Yes/No	Number of total usable addresses	**20 %**

3.2 Links with data bias and volatility

The quality of the survey data is the main empirical topic studied in the framework of the Task Force. The focus is on the volatility of the survey data and on the tracking performance of the Consumer Confidence Indicator (CCI) with respect to statistical reference series. Both of these measures may be affected by the number of participants in the survey, i.e. by the response rate.

The survey estimate studied in this section is the CCI. It provides information on the forthcoming economic development from the consumers' viewpoint. The CCI is developed for tracking the time series of private consumption. It is calculated as an arithmetic mean of questions concerning the financial situation of the household, the general economic situation, unemployment expectations and savings. All questions used in the calculation of the CCI are forward looking, i.e. the time horizon asked is over the next 12 months. (DG ECFIN, 2013)

3.2.1 Tracking performance

In general, the survey data quality is dependent on the measurement abilities of the survey questionnaire and on the sample's representation of the target population (Groves et al., 2009). The focus of the Task Force is merely on the representation issues, which include the coverage, sampling, nonresponse and adjustment errors. Usually, the difference between a sample estimate and the true population value caused by the sampling procedure is called sampling error. The other listed representation errors are referred to non-sampling errors, from which nonresponse is an interesting factor from the viewpoint of the working group of response rates. The sampling error is further discussed in Section 3.2.2 concerning volatility.

Tracking performance is measured with the correlation coefficient between the CCI and the series of final consumption expenditures of the households and NPISHs (Non-Profit Institutions Serving Households). It is obvious, that systematic nonresponse error biases the survey results and affects the tracking performance. For example, if the respondents of the consumer survey consisted merely of retired persons, the estimates describing savings and intentions to major purchases of the whole population would be severely biased. If the reference series described the whole population, then the correlation between the reference series and the sample estimate would be weak.

The correlation coefficients between the CCI and private consumption are presented on the vertical axes in the scatter plots in Figure 3.2. The correlation coefficient in the first top left scatter plot is contemporaneous, i.e. both the CCI and the reference series data come from the same time point. In the remaining scatter plots, the correlation is calculated between future private consumption with one, two or three periods ahead and the CCI (the CCI bases on forward looking economic estimates, thus it should correlate better with future consumption).

According to Figure 3.2, the contemporaneous correlation between the CCI and the private consumption varies from about 20 to about 90 percent.[5] The red points in the figure represent countries calculating their response rate as a ratio of respondents to contacted sample persons, i.e. countries presented in Table 3.2. Their response rates should not actually be compared with the response rates of the other countries, but are presented here to give an overview about all dependencies between response rates and tracking performance.

No outstanding changes occur in the correlation coefficients, if the private consumption is measured one period ahead (the top right scatter plot in Figure 3.2). More changes occur when the private consumption is measured two or three periods ahead. For Finland and for Spain, the correlation coefficients clearly increase when the private consumption is measured further in the future (with leads two or three). In Finland, the contemporaneous correlation is of 46 percent and the correlation with the private consumption 3 periods ahead is of 53 percent. In Spain, the correlation of three periods ahead is 5 percentage points higher than the contemporaneous correlation coefficient.

For some countries, the correlation coefficients of two or three periods ahead are lower than contemporaneous coefficients. The most striking decrease is seen in the correlation coefficient of Austria, which decreases about 20 percentage points when the private consumption is measured further in the future (the contemporaneous correlation is 29 percent, the correlation of three periods ahead is 10 percent). A clear decrease is also seen in the correlation coefficients of Bulgaria (from 76 to 67 percent), France (from 60 to 54 percent), Latvia (from 91 to 82 percent) and Romania (from 82 to 71 percent). The rest of the countries have quite stable correlation coefficients independent of the way of measuring the private consumption.

The response rates of the consumer surveys are presented in the horizontal axes of the scatter plots in Figure 3.2. The linkage between the tracking performance (correlation between the private consumption and the CCI) and the response rates seems to be little higher in the bottom right scatter plot with private consumption measured three periods ahead than in the top left scatter plot of the contemporaneous private consumption. The very high response rate countries (Latvia, Lithuania, Greece and Portugal) have a very good tracking performance independent of the period used in the tracking performance calculation.

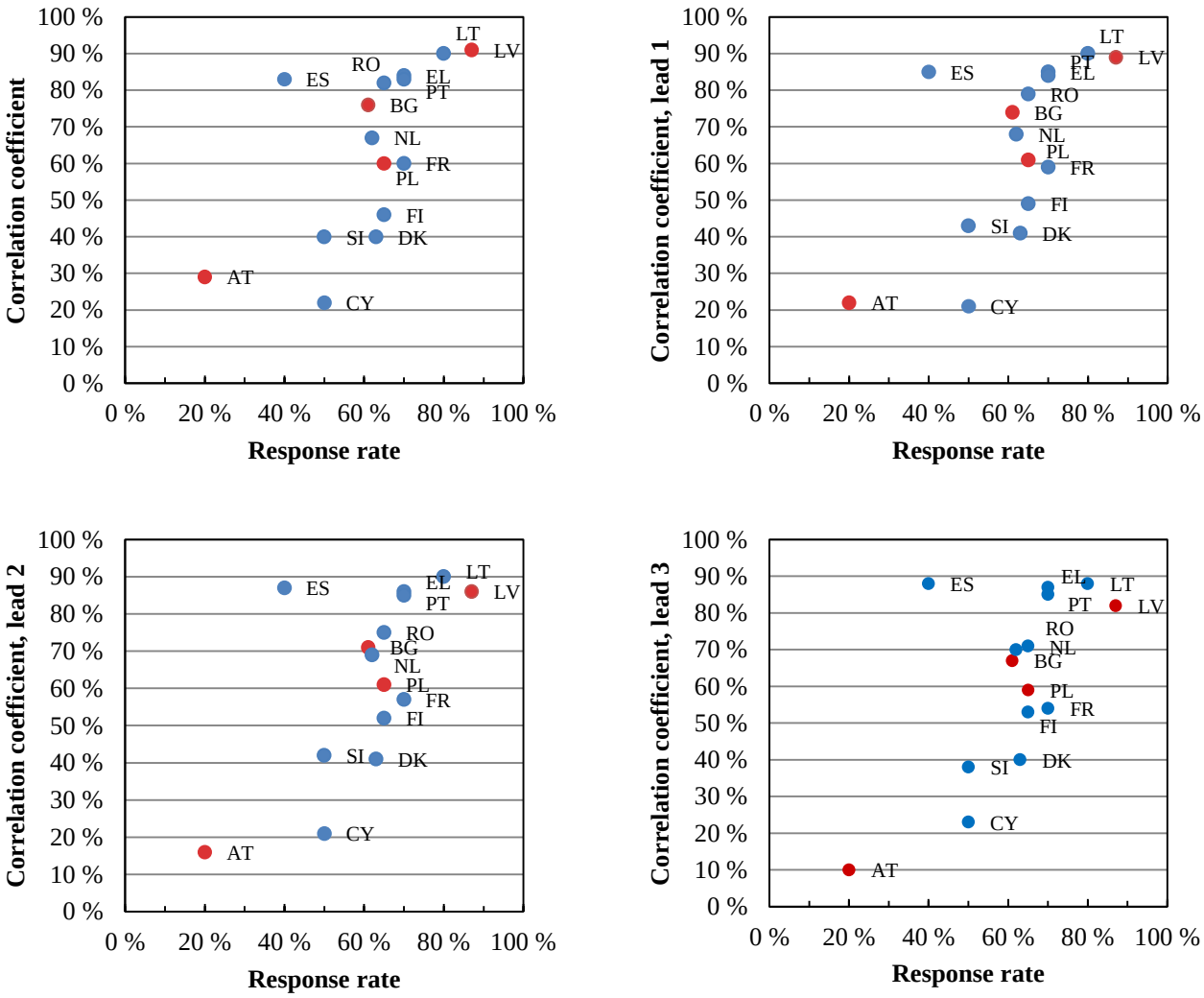
However, there are many exceptions from a clear linear trend: Spain has a low response rate but a very high tracking performance independent of the period used in the tracking performance calculation. A group of countries with response rates over 60 percent vary a lot in their tracking performance. For example, Romania and

[5] In the case of Finland, there is some evidence that CCI correlates better with durable purchases or gross domestic product (GDP) than with private consumption, but private consumption can be well predicted from household's future financial situation (Djerf & Takala 1997; Kangasniemi et al. 2010).

Denmark have both about 65 percent response rate but the tracking performance with three periods ahead in Denmark is 40 percent whereas in Romania it is much higher, about 70 percent.

The scatter plots presented in Figure 3.2 suggest that there might be some statistical dependency between tracking performance and response rates. This linkage would be interesting to study further e.g. with regression analysis but unfortunately, the quality and the magnitude of the available data do not enable proper statistical modelling. First, if the linkage between tracking performance and response rates was statistically studied, we should have the response rates calculated as an average from the complete time series. If the response rate only represents the response behaviour of a certain year, for instance, it cannot be compared with the tracking performance measure including information from a very long time series. For example, the mean response rate in Finland beginning from the year 2000 and ending on the year 2013 is about 69 percent. The mean response rate in the year 2013 was in September about 61 percent. In the metadata received from DG ECFIN the response rate in Finland was informed to be 65 percent. Secondly, at present the amount of observations available is the number of countries represented in Table 3.1 (excl. Turkey), i.e. 11 countries in all.

Figure 3.2. Dependency between tracking performance and response rate



3.2.2 Volatility

As it was mentioned in the previous Section 3.2.1, survey data quality is affected by non-sampling and sampling errors. A well-known feature of a sample survey is that because the survey estimates are based on a sample of population elements representing the whole population, sampling errors may occur. Naturally, if the data are received from a group of sample elements instead of the whole population, estimates are only an approximation of the values of population characteristics.

In statistical terms introduced by Bethlehem et al. (2011), let Y be a population parameter depending on the values in the population of one or more variables (e.g. population total or mean). The estimator of the population parameter $\hat{y}$ is the computational formula that is used to calculate the sample statistics from the selected sample. The sampling error $\hat{y} - Y$ can be measured with the standard error of an (unbiased) estimate. For example, let the unbiased estimator of a sample mean from a sample selected with simple random sampling without replacement be

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

where $y_1, \dots y_n$ denote the n observations of the population variable Y that have become available in the sample. The variance of this estimator equals

$$V(\hat{y}) = \frac{1 - \frac{n}{N}}{n} S^2,$$

where N denote the number of population elements and S^2 is the population variance. For the calculation of the standard error, an unbiased estimator is required for the population variance. The sample variance s^2 used as the estimator is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (y_i - \hat{y}).$$

The standard error measuring the sampling error is the square root of the variance, i.e.

$$S(\hat{y}) = \sqrt{V(\hat{y})} = \sqrt{\frac{1 - \frac{n}{N}}{n} s^2}. \quad (3.1)$$

From this expression, it is clear that increased sample size results in smaller sampling error. Because of the square root, a fourfold number of observations available in the sample will induce a sampling error two times smaller.

As it was mentioned in the Terms of Reference of the Task Force, in addition to reducing the reliability of the estimates, a high sampling error leads to the higher volatility of point estimates over time. If the sampling error is very high, changes from month to month become more dependent from the "noise" than from the actual changes of economic atmosphere. The dependency between the sampling error and the number of observations leads into a conclusion that the smaller the number of observations, the more likely it is that a monthly change in the estimator is due to noise rather than due to a real change in economic sentiment.

The focus of this paper has been at the response rates, which are merely related to the non-sampling error component not on the sampling error component of the total

survey error. However, with response rates we can calculate the number of observations available in the sample, which has an important role in the sampling error calculations (see equation 3.1). The sample issues are further discussed in the Task Force papers of thematic groups studying sample sizes and sampling methods. It is here assumed that these working groups study more detailed the linkage between the number of available observations and volatility. Hence, the analysis of volatility is not presented in this paper.

4 Conclusions

As a part of the Task Force launched by DG ECFIN in 2013, Statistics Finland was instructed to analyse the response rates of consumer surveys across countries participating in the EU Business and Consumer Survey Programme. Main topics investigated were factors affecting survey (non)response and the linkage between response rates and the relevance of the survey data.

The first part of this paper consisted of an overview on survey literature concerning (non)response in household surveys. The survey literature describes several things that have a connection to the number or proportion of noncontacts and refusals in surveys. On the noncontact rate of an interviewer assisted survey, the following factors are known to have an effect: number and timing of contact attempts, length of the data collection period, and workload of the interviewers. As for refusals in surveys, there are some tools that can be employed: advance and persuasion letters, respondent incentives, and survey mode. Furthermore, the role of an interviewer is an essential element for the degree of refusal to participate in household surveys.

In the empirical part of this paper we found out that the response rate of the consumer survey varies a lot between the EU countries. The minimum response rate showed up to be as low as 3 per cent and the maximum rate as high as 100 per cent. However, the computational formula for response rate is not uniform among the institutes conducting the EU consumer surveys. Anyway, one clear outcome was that survey modes with a face-to-face data collection period produce in average higher response rates than telephone surveys.

Most of the countries with the highest response rates (calculated as a ratio to the total sample size) use advance letters and have interviewer salary dependent on the number of achieved interviews. Instead, controlling the number of contact attempts surprisingly does not seem to have effect on the level of survey response in the EU consumer surveys.

As an important and interesting part of the Task Force, we studied the response rates also in context of tracking performance of the EU consumer surveys. We found out that there could be a linkage between tracking performance and response rates. Of course there were many exceptions, but the surveys of higher response rate countries appeared to have better tracking performance independent of the number of the leads. Unfortunately, the quality and the magnitude of the available data do not enable proper statistical modelling to give more support for this finding.

Finally, we suggest that a mutually agreed computational formula used in the response rate calculation would be put up to make the comparison of response rates easier. Furthermore, a more detailed analysis of response rates would also require

information about the reasons for nonresponse, i.e. noncontacts and refusals. It would be useful to see, how these rates have developed in the EU consumer surveys.

5 References

The American Association for Public Opinion Research (2011). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition. Retrieved 13.9.2013 from http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156

Bethlehem, J., Cobben, F. & Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys.* New Jersey: Wiley.

Curtin, R., Presser, S. & Singer, E. (2005). Changes in Telephone Survey Nonresponse Over the Past Quarter Century. *Public Opinion Quarterly, 69*(1), pp. 87–98.

Dillman, D.A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J. & Messer, B.L. (2009). Response Rate and Measurement Differences in Mixed-Mode Surveys Using Mail, Telephone, Interactive Voice Response (IVR) and the Internet. *Social Science Research, 38*(1), pp. 1–18.

Djerf, K. & Takala, K. (1997). Macroeconomy and Consumer Sentiment: Performance of the Finnish Consumer Barometer after Ten Years. Bank of Finland Discussion Papers 20/1997, Helsinki. Also published in Oppenländer K.H., Poser G. & Weller S. (eds.): *Selected Papers Submitted to the 23rd CIRET Conference in Helsinki,* Aldershot: Ashgate.

European Commission Directorate-General for Economic and Financial Affairs (2013). Business and Consumer Surveys. Retrieved 30.10.2013 from http://ec.europa.eu/economy_finance/db_indicators/surveys/index_en.htm

Groves, R.M. & Couper, M.P. (1998). *Nonresponse in Household Interview Surveys.* New York: Wiley.

Groves, R.M., Fowler, F.J. Jr., Couper, M.P., Lepkowski, J.M., Singer, E. & Tourangeau R. (2009). *Survey Methodology*. New Jersey: Wiley.
Groves, R.M., Singer, E. & Corning, A. (2000). Leverage-Saliency Theory of Survey Participation. *Public Opinion Quarterly, 64*(3), pp. 299–308.

Hox, J.J. & de Leeuw, E. (1994). A Comparison of Nonresponse in Mail, Telephone and Face-to-face surveys. *Quality and Quantity, 28*(4), pp. 329–344.

Junes, T. & Simpanen, M. (2013). Case Study Consumer Survey 2012: Does the Mixed-Mode Improve Response Rates and Quality of Data? Presentation presented at the fifth Conference of the European Survey Research Association (ESRA), Statistics Finland. Retrieved 3.10.2013 from http://www.europeansurveyresearch.org/conference/programme?sess=176#278

Kangasniemi, J., Kangassalo, P. & Takala, K. (2010). What Affects the Views about the Economic Sentiment? Evidence from the Consumer and Manufacturing Surveys in Finland. Paper presented at the 30th CIRET Conference, New York, October 2010.

de Leeuw, E. (2005). To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics, 21*(2), pp. 233–255.

de Leeuw, E., Callegaro, M., Hox, J., Korendijk, E. & Lensvelt-Mulders, G. (2007). The Influence of Advance Letters on Response in Telephone Surveys. *Public Opinion Quarterly, 71*(3), pp. 413–443.

de Leeuw, E. & de Heer, W. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In Groves, R. M., Dillman, D. A., Eltinge, J. L. & Little, R. J. A. (Ed.), *Survey Nonresponse* (pp. 41–54). New York: Wiley.

OECD's Glossary of Statistical Terms (2013). Months for Cyclical Dominance. Retrieved 13.9.2013 from http://stats.oecd.org/glossary/detail.asp?ID=1688

Singer, E. (2002). The Use of Incentives to Reduce Nonresponse in Household Surveys. In Groves, R. M., Dillman, D. A., Eltinge, J. L. & Little, R. J. A. (Ed.), *Survey Nonresponse* (pp. 163–178). New York: Wiley.

Singer, E., van Hoewyk, J. & Maher, M.P. (2000). Experiments with Incentives in Telephone Surveys. *Public Opinion Quarterly, 64*(2), pp. 171–188.

Stoop, I., Billiet, J., Koch, A. & Fitzgerald, R. (2010). *Improving Survey Response: Lessons learned from the European Social Survey.* Chichester: Wiley.