Chapter 1

An Introduction

1.1 Introduction

Recent years witnessed a growing demand for statistical data on the social, economic and political status of the country. Such information will enable, planners policy makers and others to make informed decisions for a better future. Sometimes, such statistical information e.g administrative record can be retrieved from existing sources. Usually, there is a lack of such sources. Therefore, a survey is a powerful instrument to collect new statistical information. A survey collects information about a well-defined population. This population is not necessarily consist of persons. A questionnaire is used collect information by asking questions to the representatives of the elements in the sample. Another way to obtain information about a population is to collect data about all its elements. Such an investigation is called a census or complete enumeration, is a very costly and time consuming approach. A survey is a solution to many of the problems of a census. Surveys collect information on only a small part of the population (sample). Sample is selected using probability sampling it is possible to make inference about the population as a whole. A random selection procedure uses an element of chance to determine which elements are selected, and which are not. Survey results allow making reliable and precise statements about the population as a whole.

Probability sampling: Is considered the best way to ensure that valid inferences can be made (Kish 1965). The theory behind sampling is well described by, amongst others, Cochran [1977] and Särndal et al (1992). It uses a randomization mechanism to give each member of the target population of interest ,a known nonzero probability of selection (Kish 1965). Probability sampling involves the concept of coverage of the target population in the sense that the survey researcher would like all units in the geographic area of interest to have a chance of falling into the selected sample. Population totals, proportions, and means can be estimated from a probability sample and the sampling variability of these estimates as measured by the standard error can also be calculated.

1.2 Types of surveys

Survey research represents one of the most important areas of measurement in applied social and economic research. There are two types of surveys cross-sectional and longitudinal. A longitudinal surveys are more complex survey design than cross-sectional surveys, it certainly offers several analytical advantages. Cross-sectional surveys are conducted at a single time point. Longitudinal surveys are relatively less expensive and take less time to conduct compared to cross-sectional surveys. The data that are collected in a cross-sectional survey may provide an opportunity to analyze many substantive outcomes, and can be helpful to achieve several objectives (e.g. public health planning). However, cross-sectional surveys ignore the fact that the same sample units may provide different measurements on the same variables if a different time frame was chosen, Thus analysis of unit-level change is not possible in these surveys. On the other hand, a longitudinal surveys are more expensive and difficult to conduct, however, they can provide data on the same set of units for a number of time points (waves). This enables the production of population cross-sectional measures every time data are collected, however, more importantly allows the analysis of unit-level change. Experts in surveys analyze the advantages and disadvantages of both longitudinal and cross-sectional surveys in different ways, however, the most recent, and probably the most informative discussion, available in (Lynn, 2009).

There are two main types of longitudinal surveys:

1- **Cohort studies** focus on a particular population, however, they sample from a specific age cohort, the sample drawn for a cohort study is selected from a birth cohort of individuals who were born in a single week or a month in a given year. The cohort study then follows the lives of the individuals selected in the sample and interviews them at particular ages at regular or (often) irregular intervals to investigate patterns in specific socio-economic phenomena. For example, a survey organization may decide to follow the lives of a sample of new born children who will be born in a single month in the year 2025 to understand the factors associated with the change in their health at different ages.

2- **Panel studies** also follow the same sample units and attempt to collect data from them at every data collection point. A major distinction between the two lies in the way they select their samples. While birth cohort studies sample from a specific age cohort, panel studies target the entire age range in a given country to investigate the dynamic of change (in a wide range of phenomena) experienced by the resident population in the country. Every year, the same individuals can be contacted and asked similar questions and the reasons for any change. Panel studies tend to have more frequent data collection points (waves) compared to cohort studies. However, this can result in extremely specific and useful explanations of social phenomena. Thus, as they target larger populations and collect data more frequently than cohort studies, panel studies tend to be more complex and more difficult to conduct. As a result, they can suffer from more problems. The focus in this research is on panel studies, in particular household panel studies, and specific type of problems (errors) that occur in these surveys which will be explained later. In recent decades, the world has seen the execution of large household panel studies. Some of these studies implement the best procedures in the art of survey design that survey research has developed. Some of the major household panel studies in the world are:

1-**The British Household Panel Survey (BHPS)** conducted in Great Britain (1991-2008). The BHPS is a result of a proposal to the UK Economic and Social Research Council (ESRC) to establish an interdisciplinary research centre at the University of Essex (Lynn, 2006). More details on the BHPS design, sample and other features will be given later as this is the main data source for this research. In 2009, Understanding Society took over from the BHPS (the BHPS was incorporated into Understanding Society) as the new UK household longitudinal study. With a sample of 100,000 individuals, Understanding Society is currently (2016) the world's largest survey of its type.

2-**The German Socio-economic Panel (GSOEP)** is the household panel study of the population in Germany. It was started in 1984, and is conducted by the German Institute for Economic Research (DIW Berlin) (Kroh, 2009).

3-**The Panel Study of Income Dynamics (PSID)** is the oldest longitudinal panel study. PSID started in the USA in 1968 and has been collecting measurements from the same sample ever since (Duncan et al, 2004).

4-**The Swiss Household Panel (SHP)** started in Switzerland in 1999. Based on the country's telephone directory, the survey covers individuals who are resident in private households in Switzerland who have a registered landline or mobile phone (Plaza, 2008).

5-**The Household, Income and Labour Dynamic in Australia (HILDA)** started in 2001 as Australia's household-based longitudinal survey. HILDA pays more attention to family and household formation, income and work than other socioeconomic topics (Summerfield, 2010).

1.3The survey process

Carrying out a survey is often a complex process that requires careful consideration and decision making. This section gives a global overview of the various steps in the process, the problems that may be encountered, and the decisions that have to be made. Figure 1.1 shows the steps in the survey process (Bethlehem2009).

The first step in the survey process is survey design. Before data collection can start, a number of important decisions have to be made. First, it has to become clear which population will be investigated (the target population). Consequently, this is the population to which the conclusions apply. Next, the general research questions must be translated into specification of population characteristics to be estimated. This specification determines the contents of the questionnaire. Furthermore, to select a proper sample, a sampling design must be defined, and the sample size must be determined such that the required accuracy of the results can be obtained.

The second step in the process is data collection. Traditionally, in many surveys paper questionnaires were used. They could be completed in face-to-face interviews: interviewers visited respondents, asked questions, and recorded the answers on (paper) forms. The quality of the collected data tended to be good. However, since face-to-face interviewing typically requires a large number of interviewers, who all may have to do much traveling, it was expensive and time-consuming. Therefore, telephone interviewing was often used as an alternative. The interviewers called the respondents from the survey agency, and thus no more traveling was necessary. However, telephone interviewing is not always feasible: only connected (or listed) people can be contacted, and the questionnaire should not be too long or too complicated. A mail survey was cheaper still: no interviewers at all were needed. Questionnaires were mailed to potential respondents with the request to return the completed forms to the survey agency. Although reminders could be sent, the persuasive power of the interviewers was lacking, and therefore response tended to be lower in this type of surveys, and so was the quality of the collected data.

Nowadays paper questionnaires are often replaced with electronic ones. Computer assisted interviewing (CAI) allows to speed up the survey process, improve the quality of the collected data, and simplify the work of the interviewers.

Computer assisted interviewing comes in three forms: computer-assisted personal interviewing (CAPI), computer-assisted telephone interviewing (CATI), and computer-assisted self-interviewing (CASI). More and more, the Internet is used for completing survey questionnaires. This is called computer-assisted web interviewing (CAWI).

If the data are collected by means of paper questionnaire forms, the completed questionnaires have to undergo extensive treatment. To produce high quality statistics, it is vital to remove any error. This step of the survey process is called data editing. Three types of errors can be distinguished:

Range error occurs if a given answer is outside the valid domain of answers (e.g. a person with an age of 348 years).

A consistency error indicates an inconsistency in the answers to a set of questions.

A routing error occurs if interviewers or respondents fail to follow the specified branch or skip instructions; that is, the route through the questionnaire is incorrect. In this case, irrelevant questions are answered, or relevant questions are left unanswered.

Detected errors have to be corrected, but this can be very difficult if it has to be done afterward, at the survey agency. In many cases, particularly for household surveys, respondents cannot be contacted again, so other ways have to be found out to solve the problem. Sometimes, it is possible to determine a reasonable approximation of a correct value by means of an imputation procedure, but in other cases an incorrect value is replaced with the special code indicating the value is unknown. After data editing, the result is a clean data file, that is, a data file in which no errors can be detected any more. However, this file is not yet ready for analysis. The collected data may not be representative of the population because the sample is affected by non-response; that is, for some elements in the sample, the required information is not obtained. If non-respondents behave differently with respect to the population characteristics to be investigated, the results will be biased. To correct for unequal selection probabilities and non-response, a weighting adjustment procedure is often carried out. Every record is assigned some weight. These weights are computed in such a way that the weighted sample distribution of characteristics such as gender, age, marital status, and region reflects the known distribution of these characteristics in the population. In the case of item non-response, that is, answers are missing on some questions, not all questions; an imputation procedure can also be carried out. Using some kind of model, an estimate for a missing value is computed and substituted in the record. Non-response problem will be discussed extensively in the next chapter as it represents the main aspect of the research problem in this thesis.

Finally, a data file is obtained that is ready for analysis. The first step in the analysis will probably nearly always be tabulation of the basic characteristics. Next, a more extensive analysis will be carried out. Depending on the nature of the study, this will take the form of an exploratory analysis or an inductive analysis. An exploratory analysis will be carried out if there are no preset ideas, and the aim is to detect possibly existing patterns, structures, and relationships in the collected data. To make inference on the population as a whole, an inductive analysis can be carried out. This can take the form of estimation of population characteristics or the testing of hypotheses that have been formulated about the population. The survey results will be published in some kind of report. On the one hand, this report must present the results of the study in a form that makes them readable for non experts in the field of survey research. On the other hand, the report must contain a sufficient amount of information for experts to establish whether the study was carried out properly and to assess the validity of the conclusions.

Carrying out a survey is a time-consuming and expensive way of collecting information. If done well, the reward is a data file full of valuable information. It is not unlikely that other researchers may want to use these data in additional analysis. This brings up the question of protecting the privacy of the participants in the survey. Is it possible to disseminate survey data sets without revealing sensitive information of individuals? Disclosure control techniques help establish disclosure risks and protect data sets against disclosing such sensitive information.

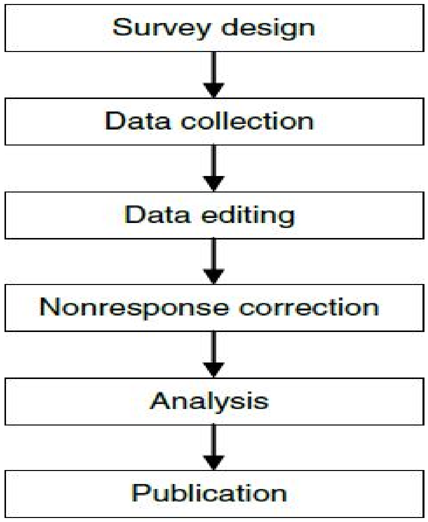

Figure 1.1: Survey process.

1.4 Survey variables

There are two types of variables that play a vital role in a survey sampling; these are survey items and auxiliary variables. According to Cobben (2009). A survey items are the variables of interest that are measured in the survey. A survey item can be the answer to one survey question, but it can also be composed of the answers to two or more survey questions. Here, a notation needs to be addressed. Every element $i = 1, 2, ..., N$ of the population U has a fixed, non random value for the survey item, denoted by Y_i. The only randomness comes from the selection of the sample. The $N \times 1-vector$ of values for the survey items is denoted by $Y = (Y_1, Y_2, ..., Y_N)'$. Auxiliary variables are in general variables that are available prior to sampling. They can be used in creating a sampling design or in the computation of the survey estimates. We denote auxiliary variables by

X_i . Chapter 3 in this thesis gives a rich background about the sources and types of auxiliary variables. Usually, more than one auxiliary variable is available in a survey. Suppose we have J auxiliary variables. Every person i can be associated with a $J \times 1-vector$ of values for the auxiliary variables $X_i = \left(X_{i1}, X_{i2}, ..., X_{ij}\right)'$. The auxiliary information for the entire population is denoted by the $N \times J - matrixX$. Usually the auxiliary variables are available for every element in the sample or the target population, for instance from a population register. When used to create a sampling design, the values of the auxiliary variables need to be known for every element in the population. However, in the estimation stage such detailed information is not always necessary. It may be sufficient to know the population total of the auxiliary variables, while the individual values are known for the respondents only. Särndal and Lundstraöm(2005) distinguish between auxiliary information on the sample level, denoted by vector X° of dimension J°, and information on the population level, denoted by vector X^{*} of dimension J^{*} . The population total $\sum_U X^{*}$ is known, for instance from an external source. However the population totals $\sum_U X^{\circ}$ is not known, but can be estimated based on the sample. In both cases, the values for the auxiliary variables are known for the responding elements in the sample. In case there is a population register, values for the auxiliary variables are available for all elements i in the population U .

Survey items and auxiliary variables can be either quantitative, qualitative or indicator variables. Quantitative variables measure quantities, amounts, sizes, or values. Qualitative variables divide the population into groups. The values denote categories.

Sample elements in the same category belong to the same group. Qualitative variables are often referred to as categorical variables. Business statistics are usually quantitative. Indicator variables measure whether or not a sample element has a certain property. They can only assume the values 0 and 1. The value is 1 if an element has the property, and otherwise it is 0. An example of an indicator variable is the response indicator. If a sample element responded to the survey, the value of the response indicator is 1 otherwise its value is 0. Survey items are identified for the target population, however we only observe them for the sample or, more precisely, for the respondents to the survey.

1.5 Errors in survey research

Requirements for sample surveys to produce the exact population characteristics are very complex and, in most cases, difficult to achieve. Thus, the estimated population characteristics are subject to error. This error can have many causes. The ultimate result of this error is a discrepancy between the survey estimate and the true population characteristic. This discrepancy is called the total survey error, [See e.g.,Seal(1962),Hansen, Hurwitz Jubine(1964),Singh(1983,1986), Bethlehem (1999)]. Bethlehem, Cobben and Schouten (2011), give a taxonomy of survey errors, displayed in figure 1.2. Two broad categories can be distinguished contributing to this total error: sampling errors and non-sampling errors. Sampling errors can be divided into estimation errors and selection errors. Estimation errors arise because only a subset of the population is surveyed. This type of error disappears if the entire population is surveyed. Selection errors occur when the actual inclusion probabilities differ from the true values of the inclusion probabilities, for instance when an element has multiple entries in the sampling frame. Consequently, sampling errors vanish when the entire population is surveyed and can be controlled for by using a correct sampling design. Non-sampling errors do not vanish when the entire population is surveyed. Non-sampling errors can be divided in errors caused by erroneous observation of sampled elements and non-observation errors. Erroneous observation arises due to overcoverage or measurement errors. In case of overcoverage, the error is caused by a discrepancy between the sampling frame and the actual population. Overcoverage occurs when sample elements that are not in the population, have an entry in the sampling frame. These sample elements are observed whereas they should not have been. A measurement error is a discrepancy between the reported or measured value and the true value. This discrepancy can be caused by the respondent, the interviewer or the questionnaire design. Measurement errors also arise when the concept implied by the survey question, and the concept that should be measured in the survey are different. Biemer and Lyberg(2003) refer to this type of error as affect the inclusion probabilities. They are caused by undercoverage or non-response. In case of undercoverage, the sample element does not have an entry in the sampling frame whereas it is part of the population. The sample element cannot be observed and the selection probability is zero. For a particular survey item Y, the set of elements that is missing consists of unit non-response and an additional set of item non-response.

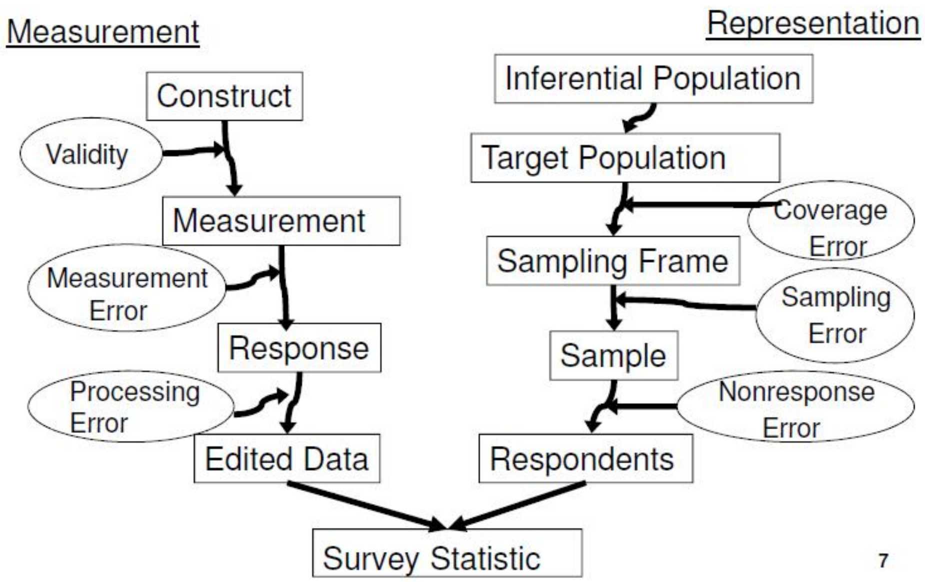

Figure 1.2: The types of survey errors.

1.6 Research problem:

The standard weighting approach (SWA) for the treatment of unit non-response assumes that non-response propensity is the same in all sub-groups in the sample.

Since, in practice, this may not be the case, SWA may lead to bias in the estimates.

A need thus arises for an approach that takes differences in non-response propensity in the subgroups into account. This motivated the present work.

Chapter 3 provides a detailed description of the research problem in this thesis.

1.7 Research Objective:

1. Explore the possibility of reducing bias in estimates propensity of unit non-response by employing an approach that takes into account differences in propensity among sub-groups in the sample.
2. Evaluate and Compare the suggested approach with the standard weighting approach (SWA).

1.8 Research Hypotheses:

This research is concerned about testing the hypothesis that

The alternative weighting approach results in the same estimates as the standard weighting approach.

Against

The alternative weighting approach results in different estimates as compared to the standard weighting approach.

Chapter 2

The Non-response Problem

2.1 Introduction

In sample surveys; "non-response is a phenomenon that affects almost all surveys, it significantly represents an origin of non-sampling error and might introduce systematic bias into data and therefore compromise the quality of the survey. Other pitfalls include: increasing the variance of the survey-based estimates and wasting part of the financial resources of the survey. Thus, it affects the derived estimates by generating poor data quality (Schafer and Graham, 2002). based on the characteristic of the survey, there are specific features of non-response that are confined to certain surveys, yet, non-response shares common features among all types of surveys. In many countries, non-response is perceived to have increased in recent years. This leads to worries about the effects of non-response on survey research.

2.2 The definition of non-response

Non-response can be defined as the phenomenon that leads elements (persons, households, companies) in the selected sample to not provide the requested data, or that the provided data is use-less (Bethlehem and Kersten (1986);Bethlehem (1988) ; Bradburn(1992);Lynn,(1996); Dillman et al.(2002);Biemer and Lyberg(2003); Voogt(2004); Bethlehem and Schouten(2004); Särndal and Lundström (2005); Bethlehem, Cobben and Schouten,(2011)). There are two types of non-response: Unit non-response is a failure to obtain any required data from a selected sample member. Item non-response, on the other hand, refers to the situation where a sample unit participates in the survey but he/she does not provide data for some of the survey questions (i.e. data for some items are not available for analysis). It is, therefore, obvious that unit non-response is more problematic than item non-response. Figure 2.1 explains the distinction between unit and item non-response from an analytical perspective.

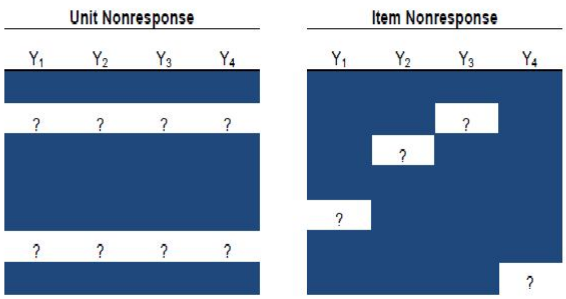

Figure 2.1: Types of non-response.

2.3 Reasons of non-response

Sadig (2011) stated that people are different in the tendency to respond to the survey. Some people make obtaining data an easy task, others are not, some people are approachable, others are not. different explanations for no-response were identified (see Lessler and Kalsbeek,1992; Lepkowski and Couper, 2002; Lynn, 2008; Groves et al, 2004; Kish, 1965), those explanations or reasons are:

(1) Non-contact: it means inability to reach sample member: for example, missing, invalid or incomplete address, even in cases in which the address is identified and contacted successfully, but the sample member is not present in that address at the time of collection (e.g. not at office or at home).

(2) Refusal: it happens when the sample member is not cooperating among contact.

(3) Incapacity or inability: the difference between this and the above “refusal” is that in this case the contacted sample member is willing but unable to participate in the survey (e.g. because of sickness, illiteracy or a language barrier).

(4) loss of information: it happens post-collection, when the collection tool (e.g. questionnaire) is lost or ruined.

A large proportion of non-responding cases are due to “non-contact” and “refusal”. To be specific, “refusal “is now the major cause (Brick, 2013; Atrostic et al, 2001).

Based on non-response correlates, researchers organize non-response into its main sources: refusal and non-contact, this will help to comprehend the factors associated with the different sources for non-response,

Refusals

Five sets of correlates of refusal were identified by (Groves and Couper, 1998). These are: Factors related to social environment; counting urbanity and crime rates in the area; factors related to respondent, including demographic features, family conformation and personal attitude; factors related to the survey like the its design, topic, sponsor, and data collection method; factors related to the interviewer such as his/her experience, age, race, gender and attitude to the survey mission; and the interviewer-respondent’s interaction as a factor. Gender

Urbanity and single-person household were found to be highly correlated with refusals, as well as survey’s topic, sponsor and design (Groves and Couper, 1998).

Gender**:** lower refusal rates were found in females; compared to males, tendency of women to participate in chats and talks is greater than men (Smith, 1984).

***Urbanity*:** refusal rate is higher in urban than non-urban areas (Steeth, 1981). Groves and Couper (1998) described it based on the fact that people of urban areas tend to avoid contact due to higher rate of crimes than rural areas.

***Single-person household*:** probably people who live by themselves are isolated socially and may not feel obligated to cooperate with surveys or to be found by the interviewer (Brehm, 1993).

***Survey topic*:** the survey topic influences individuals to cooperate according to their interest in the topic (Groves, Presser and Dipko, 2004). For example, election surveys gain more cooperation from those who are concerned in politics (Couper, 1997; Brehm, 1993). High refusal rate is associated with delicate issues (e.g. self-opinion regarding same sex marriage) (Lynn, 2008).

***Survey sponsor*:** sponsors such as governments and academic institutions generally have high cooperation rates than commercial sponsors (Groves and Couper, 1998).

***Survey design*:** designs that implement approaches to encourage sample members to cooperate yield high rate of cooperation. For example, surveys that offer reward face less refusal rates (Laurie and Lynn, 2009).

Non-contact

At-home pattern refers to non-contact due to unavailability of sample members at home at the time of the contact attempt, it happens in household surveys.

Access impediments refers to non-contact due to limiting access measures that obstruct interviewers from making the contact.

At-home pattern and Access impediments are found to be highly linked with two issues; type of sample unit and mode of data collection.

Regarding At-home pattern, for instance, people who stay away from home are difficult to contact (i.e. males, employees, and young single individuals) (Groves and Couper, 1998). In another hand, non-contact due to at-home pattern (in surveys that use call-interview) is subjected to the time of the call. Contact by calls during weekdays is less successful than that during evenings and weekends.

With regard to Access impediments, for instance, living in houses that are in locked blocks, or equipped by gates or security systems will prevent contact in face-to-face interviews (Groves and Couper, 1998). In call-interviews, answering machines and caller identity device may cause access impediments (Tuckel and O'Neill, 1995).

It is useful to analyzing non-response on the ground of “non-contact” and “refusal”, however, I believe that limiting the scope of the analysis to those two causes will show us little about the other reasons of “non-response”. It is not “non-contact” and “refusal” per se that causes non-response, other reasons include also the situations of sample members at the time that the survey invitation is made. To build a deeper insight of “non-response”, we need to go through another level and examine why it was difficult to contact those who were not contacted. For example, job hunting may keep some people outside their homes during the survey, others may not be part of the target population because they left the country and thus were not contacted. These two examples can draw into different consequences of “non-response”. Therefore, they can be useful in determining circumstances when we should worry about non-response and when we should not. Also, you can easily see that classifying “non-response” on two categories (e.g. non-contact, and not available) do not oppose each other, instead, they complement, so, these two examples can be beneficial in answering the question: when the researcher should worry about “non-response” and when he/she shouldn’t?

2.4 Effects of non-response

There are two consequences of “non-response”:

(a) Reduced sample size. Which in turn affects the precision of the estimates. However, it can be easily overridden by setting a required attained sample according to a expected non-response level (Lynn, 1996).

(b) if those who do not respond have different views from respondents on the studied variables, estimates based exclusively on information from respondents are prone to bias.

“Non-response error” refers to the increases the Mean Square Error (MSE) of survey estimates due to those two consequences that. Yet, the dominant component is the bias, and concerns are raised about non-response error (Lynn, 1996).

A deviation in a statistic that is estimated on the set of responding sample from one that estimated on a full sample is known as “Non-response bias”. This systematic distortion of the response process leads to this deviation. Considering estimates such as the mean of a target variable Y (represented by the corresponding sample statistics y) non-response error is the deviation in the value of y for respondents from the value of y for the full sample. Taking the non-respondents into account, this can be expressed by equation (2.1).

$$\overline{y_r} - \overline{y_n} = \left(\frac{m}{n}\right)\left(\overline{y_r} - \overline{y_m}\right) \qquad (2.1)$$

The selected sample size is represented by n , the number of non-respondents is denoted by m and r indicates the respondents. Thus, y_r is the value of y for respondents; y_n is the value of y_r for the full sample; and y_m is the value of y for non-respondents.

2.5 Longitudinal survey and Non-response

Lately, longitudinal surveys have gained more attention. leading to enhancement in their design and implementation. For instance, a decrease in the cost of surveys and enhancement of the quality of the data has resulted from the including computer technology in data collection (Bethlehem, 2009). Nevertheless, a rise in "non-response" rates is considered as newly generated problems. "Non-response" rates are still increasing in most longitudinal surveys all around the world in spite of the effort invested by survey organizations to enhance the survey design (Watson and Wooden, 2009). Implementing post-survey adjustments is becoming more common. Gathering of observations from individuals on multiple times is a characteristic of household longitudinal surveys, it allows follow up with individuals over time.

Yet, joining the survey every time data is collected may not be always possible because the respondents my not be accessible. Which in turn results in "non-response" due to "non-contact" or "refusal" (Lepkowski & Couper, 2002). For instance, non-response can result when the sample member changes his/her address without telling the survey organization; also, other respondents who participated previously may refuse to respond at some point. Therefore, non-response is dynamic phenomenon in longitudinal that can happen over and over (Watson and Wooden, 2009). If the respondent is not replaceable (e.g. selected through a probability sample design); this will make non-response a problem. The challenge is to come with answers for the following questions: if a probability sample would suffer from high non-response over time, would it have an advantage? Would it be a better choice considering that it may not be representative of the target population.

In longitudinal surveys, “non-response” can be in one of two forms:

(1) When a sample member is inaccessible from the survey for at least one wave but becomes available in a later wave, this is known as “Wave non-response”

(2) When a sample member cease to continue in the following survey waves although he/she has participated successfully in the first waves, this is attrition, this has attracted more concern from survey researchers, because first: attrition leads to more loss of information, second,

It is more likely to get bias by non-response (Chang, 2010).and third, as more waves are established, the information gathered in the past waves become less reliable. (Chang, 2010). It is always hard to differentiate “wave non-response” from “attrition” in surveys with an unlimited number of waves, because it is subjected to the behavior of the respondent in the future.

2.6 Sources of “wave non-response” and “attrition”

“Wave non-response” and “attrition” happen in panel studies as special e of “non-response”. There are two differences between longitudinal surveys and cross-sectional surveys with regard to the reasons of “non-response”:

(a) Respondents are loaded by a continuous long-term obligation to responding in the longitudinal survey.

(b) Alterations in the gaps of data collection points may affect the response process. Thus, some reasons for non-response may be confined to longitudinal studies. For examples:

(1) Contact information not updated: happens when a participant in the survey change his/her address in-between waves without notifying the organizers of the survey, which makes it could track him/her cost effective and lead to “non-contact”.

(2) Loss of interest: the job of the survey organizers is to preserve the commitment of some respondents to continue their participation at every wave (e.g. use of incentive). Inability to do so leads to “refusal”.

(3) Alterations in health status: because longitudinal surveys are managed over a long time. during this, some respondents could suffer from deteriorated health status causing dropouts.

(4) Technicalities associated with data collection approach: Survey organizers may face situation that make them implement changes in the data collection approach. For instance, inability to provide the interviewer from the last wave may affect the tendency to collaborate in the current wave. This is mostly common when introducing junior interviewer who has less experience. While there is no clear evidence that the changing the interviewer leads to “non-response”, recent study shows that preserving the same interviewer is linked to low tendency for “refusal” (Lynn, Kaminska and Goldstein, 2014).

2.7 The effect of non-response in the longitudinal context

In longitudinal surveys, non-response can cause bias and diminished sample size, this is turn leads to complications:

(1) following continuous waves, the sample size might end up smaller relatively than the original size at start, this smaller sample is not suitable for creating precise estimates. This is not the same for cross-sectional surveys since we can set the required reached sample giving a predicted non-response level. However, regarding longitudinal surveys, even if we put an effort in a very large sample size, the decrease of the sample size with more waves in the long term can still be a dilemma, especially in surveys of unlimited length. This difference demonstrates that the reduction in the sample size due to non-response is more challenging in longitudinal surveys, and it is a typical character of panel data. Thus, it’s important to construct approaches and strategies to overcome this problem, to increase the size of the achieved sample and eventually improve the precision of the estimates.

(2) It is possible that dropping out of the study is not random (Fitzgerald, Gottschalk and Mofitt, 1998; Watson, 2003) probably some of the drop outs are due to the issues investigated by the survey. For instance, the respondents may choose to drop out after realizing the topic of the survey, although he/she have successfully participated in one or more waves at the beginning., as more waves are established, the sample becomes increasingly unrepresentative of the population considering that the drop outs are different from respondents in terms of what the survey is measuring. Consequently, bias will be introduced to the estimates.

Chapter 3

Treatment of non-response

3.1 Introduction

According to Bethlehem and Schouten (2004), it is very difficult, in practice to assess the possible negative effects of non-response. And even if such effects can be detected, it is no simple matter to correct for them. Vital for useful detection and correction techniques is the availability of at least some information about the non-respondents. Such information is usually available in the form of auxiliary variable . Section 3.4.5 in this chapter gives a rich background about the types and sources of auxiliary variables.

Sadig (2011) states that "there are generally two approaches to deal with non-response: dealing with non-response at the data collection stage and dealing with non-response at the analysis stage". In what follows, explain these approaches and the best practices in survey research that are involved.

3.2 Dealing with non-response at the data collection stage

Here efforts are focused on minimizing non-response when collecting the data (Lynn, 1996; Stoop et al, 2010). And in this regard, survey organizations may incorporate a mixture of techniques in the survey design in their attempt to decrease non-response rate to its minimum. Groves et al (2004) provide a wide range of these. Examples of design features that may reduce non-response are:

1- ***Increased number of contact attempts*:**

Several studies (e.g Goyder, 1985) showed that increase in the number of contact attempts usually leads to increase in the number of successful contacts.

2- ***Long data collection period*:**

The longer the data collection period the more it becomes possible to deliver the survey request to a larger group of sampling units.

3- ***Reduced number of sample members assigned to interviewers*:** The smaller the number of cases assigned to an interviewer the more is the effort given to convince some cases to respond.

4- ***Pre-notification letters*:**

Unexpected visits from interviewers may cause some sample members to refuse to respond.

5- ***Use of incentives:***

Offering financial incentives encourages response. (Laurie and Lynn, 2009).

6- ***Mixed-modes of data collection*:**

Since not all sample members can be contacted with same mode; a mixture mode design may enable contacting a larger number of sample members.

7- ***Interviewer/ household matching*:**

Assigning interviewer with characteristic acceptable to sample member increases trust & hence the likelihood of response.

3.3 Dealing with non-response at the analysis stage

The techniques referred to in section (3.2) are useful for dealing with sources of non-response at the collection stage. Other sources of non-response can be dealt with at the analysis stage. Methods used at this stage are based on the idea that the distribution of the responding sample should be adjusted to make it similar to that of the population. The adjustment compensates for those missing from the sample leading to less bias in the estimates from it. These methods are called post- survey adjustments. Before introducing the different types of adjustments that are used to in the analysis stage, it is important to distinguish between different types of missing data mechanisms and non-response patterns.

3.3.1 Missing data mechanisms

A missing data mechanism is the process that generates the missingness.

Let Y be a target variable for which some observations are missing, and X is a set of auxiliary variables that are fully observed for both respondents and non-respondents. Also Z is an outside phenomenon that is not related to Y and X, and R is an indicator variable represent the missing data state (i.e. R=1 if Y is observed, and R=0 otherwise). The mechanism for missing data can be written formallyas:

Let Y be divided into an observed part Y_{obs} and a missing part Y_{miss} . So Y = (Y_{obs} , Y_{obs}).

The distribution of the missingness is characterized by the conditional distribution of R given Y:

$$P\left(R \mid Y\right) = P\left(R \mid Y_{obs}, Y_{mis}\right) \tag{3.1}$$

Three categories of missing data mechanisims can be distinguished (Rubin1976):

1. The conditional distribution of R given Y does not depend on the data at all. This category is called missing completely at random (MCAR). In this case the mechanism that generates the missing values is a truly random process unrelated to any measured or unmeasured characteristic of the sample members. In other words, the missingness is

caused by the outside cause Z. In this case, the probability of a value being missing is unrelated to the data on that unit. Formally this can be put as:

$$P\left(R \mid Y\right) = P\left(R\right) \tag{3.2}$$

2. The conditional distribution of R given Y depends on the observed data, but not on the missing data. This is called missing at random (MAR).We say data are MAR if given, the observed data the probability distribution of R is independent of the unobserved data. Formally:

$$P\left(R \mid Y\right) = P\left(R \mid Y_{obs}\right).\ \ So\ P\left(R \mid Y_{obs}, Y_{mis}\right) \tag{3.3}$$

The causes for missing data are identified by the observed data and this means we have an opportunity to use X to adjust for this missiness.

3. The conditional distribution of R given Y depends on both the observed and missing data. This situation is called not missing at random (NMAR). Here the missingness process depends on unobserved and observed measurements. Under a NMAR assumptions, the probability of an observation being missing depends on the underlying value, and this dependence remains even given the observed data, which means that the causes for missing data are not fully identified by the observed data. This implies that:

$$P\left(R \mid Y\right) = P\left(R \mid Y_{obs}, Y_{mis}\right) \tag{3.4}$$

The three missing data mechanisms are shown in Figure 3.1. As the figure shows, the case of MCAR is brought on by a phenomenon Z that is completely unrelated to X and Y. In this case non-response has a minor effect; as it only reduces the sample size. Estimates for parameters of Y will not be biased. The case of MAR is caused partly by an independent phenomenon Z and partly by an auxiliary variable X leading to is an indirect relationship between Y and R. As a result estimates for Y are biased. Fortunately, it is possible to correct for such a bias by using a technique that takes advantage of the availability of all values of X, both for respondents and non-respondents. One has to assume the relation between X and Y can be identified by just using the observed data. The case of NMAR suggests a relationship between Z and R and between X and R. But there is also a direct relationship between Y and R that cannot be accounted for by X. This situation leads to biased estimates for Y. Unfortunately, correction techniques using X fails to remove the bias.

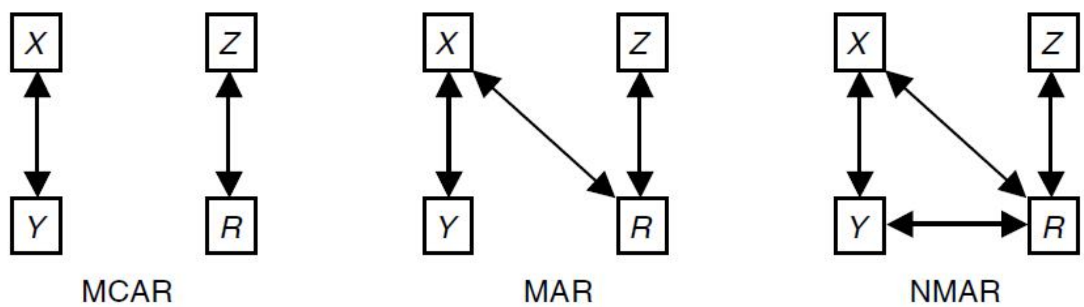

Figure 3.1: Missing data mechanism

3.3.2 Non-response patterns

Missing data patterns describe which values are observed and which ones are missing (Cobben(2009) and Allison(2002).). Different treatments are needed for different patterns of missing data. Here we consider a number of possible ways that missing data can take. We focus on the relationships between the missing values and the recorded values in the data sets.

Most survey data sets can be presented in rectangular or matrix form, so that the rows correspond to the observational elements and the columns correspond to the variables. Schafer and Graham (2002) identify three non-response patterns as shown in Figure3.2.

Figure 3.2a shows the case of a univariate missing data. The auxiliary variables $X_1, X_2, ..., X_p$ are observed for all sample elements. The column Y represents a group consisting of one or more target variables of the survey. The values of these variables are either completely observed for all variables or missing for all variables.

Figure 3.2b shows the case of a monotone missing data pattern. The target variables (or groups of target variables), $Y_1, Y_2, ..., Y_p$ may be ordered in such a way that if Y_j is missing for an element, then $Y_{j+1}, Y_{j+2}, ..., Y_p$ are missing as well. Figure 3.2c shows the case of an arbitrary missing data pattern. There is no structure or ordering in the missingness of the values of the target variables.

The term unit non-response refers to the case where data are missing according to a univariate pattern. All values of variables corresponding to questions on the questionnaire form are therefore missing. Variables obtained from registers or sampling frames may be added to the survey data file and for these variables, all values are available. The monotone non-response pattern can arise in longitudinal studies with attrition, where subjects drop out before to the end of the study and do not return. Missing values due to item non-response have an arbitrary pattern. This means that any set of variables may be missing for any unit. Note that it is possible for all three patterns in Fig(3.2) to occur in one survey.

The missing data indicator R is used to show which variable values are available and which are missing. The form of R depends on the missing data pattern. For univariate missing data, R would be a binary variable that for each sample element i takes the value 1 ie $R_i = 1$ if data are observed and 0, $R_i = 0$ if missing . For the monotone missing data pattern, R would be an integer variable (assuming one of the values $1, 2, ..., p$, where p is the number of [groups of] variables. Here $R_i = j$ means that values of $Y_1, Y_2, \ldots, Y_j$ are observed; the values of $Y_{j+1}, Y_{j+2}, \ldots, Y_p$ are not observed. For arbitrary missing data, R would be a matrix of binary indicators of the same dimension as the data matrix, with elements of R set to 1 or 0 according to whether the corresponding data values are observed or missing. An excellent discussion of these patterns and others is given in Little and Rubin(1987,2002) and Schafer and Graham (2002).

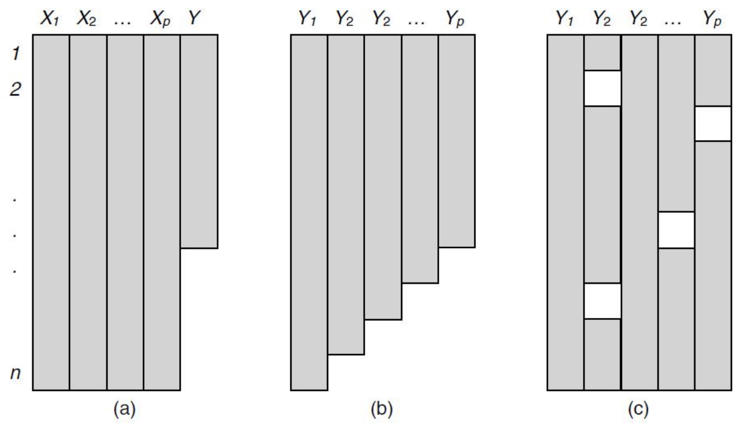

Figure 3.2: Non-response Patterns

3.3.3 Post-survey adjustments

The term "post-survey adjustments" is used to refer to methods that are used to treat errors of non-response. Several methods are available to deal with the non-response problem. The most common methods are: Non-response Weighting; Post-stratification; Calibration; Raking; Multiple Imputation; and the Selection Model Approach.

- **Non-response weighting**:

In general the term "weighting", means that every responding object in the survey is assigned a weight, and estimates of population characteristics are obtained by using weighted observations instead of the observations themselves. Non-response weighting is based on the use of auxiliary variable. Those are variables that are measured in the survey, and for which information on the population distribution is available. By comparing the population distribution of an auxiliary variable with the sample distribution, non-response error can be assessed. If these distributions differ considerably, one may conclude that non-response has resulted in a selective sample. The auxiliary variable can also be used to compute adjustment weights. Which are then assigned to all records of responding elements. Estimates of population characteristics can then be obtained by using the weighted values instead of the unweighted values.

The weights are defined in such a way that population characteristics for the auxiliary variables can be computed without error. So when weights are applied to estimate population means of auxiliary variables, the estimates must be equal to the true means. If it is possible to make the sample representative with respect to several auxiliary variables, and if these variables have strong relationship with the phenomena to be investigated, the (weighted) sample will also be (approximately) representative with respect to these phenomena, and hence estimates of population characteristics will be more accurate.

Weighting for non-response is a technique that assigns numerical values (weights) to the responding sample units, in order to modify them to also represent non-responding sample units (Lynn, 2005). As a result, it is hoped that the weighted distribution of the responding sample will be similar to that of the selected sample. More details on non-response weighting, including the construction of the weights, will be discussed in the next sections as this is the subject of this thesis. However, the term weighting will be used here as shorthand for non-response weighting.

- **Calibration**:

This is a method that assigns values (also called weights) to respondents so that known parameters of the auxiliary variables X, either from the survey or another survey, can be reproduced (Sikkel, Hox and de Leeuw, 2009; Särndal and Lundström, 2005). This procedure usually results in estimates with smaller standard errors. If the auxiliary variables used in calibration distinguish response from non-response (i.e. are correlated with the response propensity), non-response error can also be reduced.

- **Post-stratification**:

This also assigns values to respondents so that their sums are equal to known population totals for certain sub-groups of the population (Biemer and Christ, 2008), post-stratification assigns weights to respondents so that their distribution by the defined subgroups is the same as the known population distribution. In this respect, post-stratification can also be classified as a calibration method. The difference is that in calibration the known subgroups totals may not necessarily be from the population, under study: Although post-stratification is primarily used to correct for non-coverage error, and for reducing the variance of survey estimates, yet, if the auxiliary variables that form the subgroups in the post-stratification are powerful predictors of the response probability, post-stratification may also reduce non-response bias.

- **Raking**:

Raking is an extension to post-stratification that performs multidimensional post-stratification. It assigns values to respondents in order to match known population distributions in a number of auxiliary variables. Raking repeats this process a number of times until an accepted distributions are met. It, thus, differs from post-stratification in that it does not reproduce the exact population distributions on the auxiliary variables. Another difference between raking and post-stratification is that, the joint distribution of the auxiliary variables need not be known. Actually, raking can be used, if only the marginal distributions of two or more auxiliary variables are known.

- **Multiple imputation(MI):**

Is different from single imputation (SI) in that the lat-ter produces one synthetic value to replace a missing value in a target variable Y. This can be deterministic if, for example, the missing values in a variable replaced by the mean value of the variable; and random if the imputed values are selected randomly from the available values of the variable being imputed . Bethlehem (2009) provides a extensive discussion for a range of different SI methods. These are not discussed here as our focus is on MI. However, two major disadvantages of SI indicated by de Leeuw et al, (2003) that can be mentioned:

1- using the observed data to impute the missing values emphasizes the structure of the observed data in the imputed data set.

2- analyzing the imputed data set involves using a spuriously large number of cases which may lead to biased significance tests.

MI, on the other hand, may solve the problems of SI. MI produces a set of synthetic values to replace a missing value. The method originated in application to non-response (Rubin, 1976). The ordinary concept of MI proposed by Rubin,(1987) is based on three steps:

First impute the missing values in the data m times resulting in m complete data sets; Next perform the desired analysis on each of the complete imputed data sets; and finally combine the results obtained from the m-time repeated analysis into a single result. Although analyzing the data m times may seem inconvenient, yet it is not difficult, especially with the presence of a number of powerful softwares. What might be difficult is the generation of m data sets in an appropriate manner (de Leeuw et al, 2003). In MI, the imputed values must include an error term from an appropriate distribution (generally, the models used for the data generation should include variables that predict either the missingness or the outcome variable). This solves the problem of emphasizing the existing structure in the data. Also, analyzing m data sets and combining the resultant estimates into an overall estimate resolves the problem of the biased significance tests. Several developments led to MI imputation models which use variables that predict both the missingness and the outcome of interest (Schafer, 1997) which results in a more efficient analysis. Recent studies have demonstrated that MI can also be incorporated in dealing with non-response in substantive longitudinal data analyses (Goldstein, 2009; Carpenter and Plewis, 2011; Plewis, 2011). A basic difference between imputation and calibration-based methods is that while imputation attempts to produce a distribution that resembles the true distribution of the imputed variable, this is not required by calibration-based methods. All of the above post-survey adjustments assume MAR. In circumstances where survey researchers have reasons to believe that MAR does not hold, the missing data mechanism is not ignorable, and valid estimation may require modeling the missingness as part of the estimation process. MI can produce valid estimates, in this case, if the model for missingness is correctly specified (Allison, 2000). However, these situations are fraught with difficulty. Because the very data that suffer from missingness cannot support the specification of an appropriate model that correctly predict the missingness.

- **Selection model approach (SMA):**

This is a method that assumes NMAR. As stated earlier, if data are not missing at random, there are no simple solutions and a specific model for missingness must be hypothesized. The SMA postulates a model that links the missingness to the distribution of the outcome variable (Heckman, 1979; Hausman and Wise, 1979). However, SMA suffers from several weakness. A part from the fact that there is no information in the data to help chose an appropriate model, there is no statistics that can show how well a chosen model fits the data and the results are often sensitive to the choice of the model (Little and Rubin, 2002). Furthermore, applying this method requires the availability of variables that are not correlated with substantive outcome. Fully observed instrumental variables (from previous waves in case of longitudinal data) that vary between units and predict the missingness may be good used. However, such variables, may be, difficult to find. Also, other variables such as characteristics of interviewers and interview condition have little variation across units (Fitzgerald, Gottschalk and Mofftt, 1998). As mentioned earlier, adjusting for missing data appropriately depends on the missingness mechanism and the method that is used as a post-survey adjustment. However, simulation study by Collins, the MAR-based methods result in little bias in estimates even when the missingness is NMAR. The only exception is that when some of the causes of missingness that are not included in the adjustment are strongly correlated with the substantive variable Y (with a correlation coefficient greater than 0.4).

3.4 Weighting

Weighting is an adjustment which is implemented at the stage of analyzing the data. It is used to compensate for the missing units from the selected sample (non-responders). It adjusts the responding sample so that its distribution is the same as the selected sample, and hence produces unbiased estimates. In weighting we calculate numerical entities (weights) that represent the influence of survey respondents on estimates. When constructing a survey-based estimate, weighting assigns the calculated weights to respondents as their contribution to the estimate in question (Lynn, 2005). The weight of a respondent can be interpreted as the number of individuals in the target population that are represented by the respondent. Aside from non-response weighting and calibration-based weights, there is another type of weighting that is usually used in conjunction with non-response weighting. This weighting is used to reflect the selection procedure, when the sample is selected with unequal probabilities of selection. Weight values that result from this type of weighting are referred to as Design weights. In practice, the design weights are created first before adjusting with non-response weights. In this respect, the final analysis weight used to adjust for non-response is a combination of the design weight and non-response weight.

3.4.1 Design weights

The design weights are used to correct for the unequal probabilities of selection. This occurs when some of the units in the sampling frame have a different chance of being selected than other units. If a sample is selected with unequal probabilities, estimates such as the unweighted sample mean will be biased (Horvitz and Tompson, 1952). For example, consider a sample design that aims at randomly selecting one adult from each of H households. In this example the chance of an adult being selected from household h depends on the number of adults in this household. In other words, the probability of selection increases as the number of adults in the household decreases. Thus, ignoring the fact that the selection probabilities are different will result in bias in estimates due to an over-representation of households with fewer adults. This can be avoided if a correction is implemented to balance the probabilities of selection. This correction is the design weight: it adds more value to the cases whose probability of selection is low to represent more cases of their category, and decreases the value of the cases whose probability of selection is high, in order to balance the sample. Therefore, the design weight for a given unit in the sample is the inverse of the selection probability for this unit. Thus, calculating the design weights only requires knowledge of the selection probabilities for every unit in the sample. The design weight is given by:

$$D_i = P_i^{-1} \tag{3.5}$$

Where D_i is the design weight for case i; and P_i is the probability of selection for unit i. If sample units were selected using a simple random sampling method, p_i becomes constant. In this case, all sample units will have the same design weight which is the ratio of the number of units on the sampling frame to the number of units in the selected sample. Otherwise, the design weight must reflect the strategy of selection for each unit separately.

3.4.2 Constructing non-response weights

Although the rationale behind non-response weights is obvious, there is no universally held method for computation. Weights construction varies according to the differences in circumstances from sample to sample concerning the design and the availability of auxiliary information about the sample and the target population (GATS Sample Weights Manual, 2009). Thus, the actual stages for deriving the weights may vary from one survey to another. Therefore, the weights are usually created and released by the survey organization. Nevertheless, there are general well-known steps for constructing the weights, to compensate for non-response. These shall be discussed here.

Non-response weights are based on the response propensity which is measured by the probability of response. Those whose characteristics lead to low response probability should have high weight values to represent more individuals from their category, since they are less likely to respond. In turn, individuals with characteristics that lead to high response probability should have low weight values to represent fewer individuals from their category, since they are more likely to respond. Thus, a non-response weight is basically the inverse of the response probability (propensity). This is why part of the literature in this area refers to non-response weighting as Inverse Probability Weighting (IPW). There are two ways to estimate the response probability for units in the sample in order to calculate the weights: weighting classes; and model-based methods.

Weighting classes

This method is a simple approach that involves dividing the sample into a number of non-overlapping sub-groups using a few auxiliary variables (also called weighting variables) that are known for both respondents and non-respondents (Kalton and Flores-Cervantes, 2003; Little, 1986; Brick, 2013 and Biemer and Christ, 2008). The resultant sub-groups are referred to as 'weighting classes'. The response probability for each weighting class is then calculated as the class response rate and the non-response weight that is assigned to a responding unit in it is simply the inverse of the response probability of the class to which the unit belongs. For example, for a given class c, if the number of units is denoted by n_C, and the number of responding units is denoted by mc, the response probability is defined by $\frac{m_c}{n_c}$.Thus, the weight of a responding unit in class c is $\frac{n_c}{m_c}$. If there is homogeneity in terms of response propensity between all units in a weighting class (i.e. all units have the same response propensity), MAR holds, and non-response bias will be eliminated by using the weights. An alternative term which is sometimes used for weighting classes in the literature, is Response Homogeneity Groups (RHGs),see for example (Brick, 2013).The disadvantage of this method is that classes are subjectively identified in one or two dimensions, by using one or two auxiliary variables. Also, classes with small number of respondents produce small response rates and, hence large weights. Larger values of weights may introduce large variances in estimates. Lynn (1996) suggests avoiding weighting classes with a response propensity that is less than one-fifth of the overall survey response rate.

Model-based methods

In this method usually a binary outcome regression model is used to estimate the response propensities for the sample units. This method was incorporated into the survey non-response problem by David et al (1983). It is an extension of the propensity score theory of Rosenbaum and Rubin (1983). Models used in this regard, often, are referred to as Response Propensity Models (RPMs). With a suitable function, usually logit or probit, the probability of response can be medaled (response =1; and non-response =0). The non-response weights can then be calculated, for responding units, as the inverse of the predicted values from the model. For example, if R_i denotes the outcome variable in a RPM that uses a logit function (i.e. logistic regression), R_i is an indicator with the following values:

$$R_i = \begin{cases} 1 & ; \quad \textit{if the } i^{th} \textit{ sample unit responds} \\ 0 & ; \quad \textit{otherwise} \end{cases} \tag{3.6}$$

Auxiliary variables (or weighting variables) that are available for both respondents and non-respondents, and are thought to be correlated with R_i can then be used to estimate the model.

For responding units, non-response weights are then computed as:

$$W_{NR_i} = r_i^{-1} \tag{3.7}$$

Where W_{NR_i} is non-response weight for unit i ; and r_i^{-1} is the inverse of the predicted value of R_i . Using a RPM to estimate the response probability for sample members may be more effective than applying the weighting classes approach. This is because a large mixture of dummy and continuous weighting variables can be used to fit a range of models, and therefore obtain more effective non-response adjustments. However, an important disadvantage is that the predicted response probabilities for some units may differ considerably. This may result in large weights variance. Large weights variance, in turn, will increase the variance of estimates. Nonetheless, the estimated response probabilities can be grouped into weighting classes, and weights can then be recalculated using either the mean predicted probability in the class or the observed response rate in the class. Since we always assume that being selected in the sample is independent of responding to the survey, the weight that is usually used in the analysis is constructed as the product of the design weight and non-response weight. This way, every unit in the sample is adjusted using its chance of being selected in the sample and its tendency to respond to the survey simultaneously. The final analysis weight assigned to the i^{th} responding case is given by:

$$W_i = D_i * W_{NR_i} \tag{3.8}$$

Where W_i is the final analysis weight; D_i is the design weight; and W_{NR_i} is the non-response weight.

3.4.3 Effect of weighting

Weighting is used to reduce bias in survey estimates. The underlying assumption for this is that characteristics of respondents in a weighting class (or with a given set of characteristics of the auxiliary variables that predict the probability of response) are similar to the unobserved characteristics of non-respondents in the same class with respect to the survey target variables (Lynn, 2005). When this assumption is met, weighting will then successfully reduce bias from estimates. However, there is a drawback to weighting. That is variability in the weights will increase the variance of the survey estimates. Thus, while un-weighted estimates may be biased but more precise, weighted estimates are less biased but also less precise. This is an inevitable trade-off to be made in weighting. However, to limit the extent of the increase in variance, survey researchers sometimes restrict large weights to some arbitrary maximum value at which they can tolerate its corresponding increase in variance. This technique is referred to as trimming.

3.4.4 Why use weighting?

Each of the post-survey adjustments has its advantages and disadvantages. They share similarities such as requiring auxiliary variables, but they differ in terms of the way they handle non-response. Weighting and calibration-based methods assign weights to respondents as compensation for those who are missing; whereas imputation methods attempt to estimate the missing values in the substantive variables. This raises the question as to whether weighting-based methods have advantages over imputation-based methods or vice versa. To give an insight into this issue, this section compares weighting with Multiple Imputation (MI) and the Selection Model approach (SMA). Weighting relies on the MAR assumption. SMA works on the basis of NMAR. MI could be used under both MAR and NMAR, but the latter may require MI to correctly specify the model for missingness. MI views both unit and item non-response as a missing data problem. Consequently, it corrects simultaneously for unit and item non-response. Weighting, on the other hand, can only deal with unit non-response. Also, weighting ignores the association between the auxiliary variables and the outcome variable, which may lead to inefficiencies in the analysis (Plewis, 2011). Meanwhile MI and SMA take the association between the auxiliary variables and the outcome variable into account by establishing models that liThus, the estimation of different substantive models may need the application of different MI and SMA models. Weighting, however, is multipurpose.

Once the weights are created they can be used in the estimation of different substantive models (i.e. the same set of weights is used every time). In addition, for secondary data users, who are concerned about the effect of non-response on their estimates, and who do not have the technical capabilities nor the necessary data to perform a procedure like MI or SMA, weighting may be a good option. It is relatively easy to use weights in most statistics software. The weights are in the form of a variable in the data set. Usually, users only notify the software that they would like to implement weighting in their analysis and simply identify the weighting variable. In return, the software carries out the necessary calculations and produces weighted survey estimates. Moreover, analysts are able to use many standard analysis techniques with weights.

3.4.5 Non-response weighting variables

As mentioned earlier, post-survey adjustments rely on auxiliary variables in their treatment of non-response. For decades, the term auxiliary variables was mainly used to describe variables that are not of analytical interest. In cross-sectional surveys, such variables are typically available from the sampling frame from which the sample was drawn. Also, they may be available from sources external to the survey, for example, from a national census. Therefore, auxiliary variables may be available for the full sample. In this sense, auxiliary variables are by definition not of substantive interest to the survey, as designing a survey to collect variables that already exist is unnecessary. However, when longitudinal surveys emerged, they provided the opportunity to use substantive variables that were collected in earlier waves as auxiliary variables to adjust for missingness in later waves. Auxiliary variables that are used in weighting adjustment in particular are sometimes referred to as weighting variables (see for example Kreuter and Olson, 2011). In this research, regardless of the type of these auxiliary variables, we present them with the label weighting variables. The choice of weighting variables plays an important role in reducing non-response bias. In recent years, survey researchers have laid the foundation for principles to guide the selection of the best set of variables to adjust for non-response (Särndal and Lundström, 2005; Little and Vartivarian, 2003; 2005). A variable is said to be powerful in reducing non-response bias if: it shows evidence of explaining the response propensity, it is highly correlated with the survey main variables, and it identifies or comes close to identifying one of the important domains in the population. Little and Vartivarian (2005) demonstrated that if the association between the weighting variables and the variable of interest is low, the weighted mean will have increased variance without decreasing the bias even if the association between the weighting variables and the response propensity is high. In sum, in order for non-response weights to be effective in reducing bias, the weighting variables have to be correlated with the variable of interest and the response propensity. Furthermore, to be able to create the weights, the weighting variables have to be observed for both respondents and non-respondents. However, even with fewer restrictions, the existence of a good set of weighting variables, in practice, may be rare. This is because, first, in practice, only a few variables are available for both respondents and non-respondents. This is why, in recent years, survey researchers have extensively investigated alternative sources of variables that can be observed for all sample members, and advised survey organizations to move towards data collection modes that collect such variables. Second, even within the available variables, any given variable is likely to differ in the strength of its correlations with the substantive survey variables (Kreuter and Olson, 2011). Third, no single variable is likely to predict the response propensity and be correlated with all substantive variables simultaneously (Kreuter, Lemay and Casas-Cordero, 2007; Groves, Wagner and Peytcheva, 2007; Kreuter and Olson, 2011).This is why survey organizations should have plans to identify sources of potentially good variables and collect them at the data collection stage. Weighting variables can be drawn from multiple sources. These sources could be internal or external to the survey. Depending on the type of variables and the source, the main categories of weighting variables are:

(a) variables about the process in which the survey data were collected. This type of variables is referred to as paradata; for example, what was the mode of data collection (phone, web, mail, or in person).

(b)variables based on the interviewers observations about some characteristics related to the household/individual (e.g. type of accommodation).

(c) variables taken from the sampling frame, i.e. traditional auxiliary variables. These are usually available if the sample is taken from administrative records (e.g. levels of proficiency or education).

(d) variables linked from another database. Sometimes the sampling frame does not provide much information about sample units, for example, if the sample frame is the postcode address file (Lynn, 1996). In this example, although the postcode itself does not provide information about sample members living at the selected address, it can be used to link geographical information from another database such as credit scores (Lynn, 1996).

(e) substantive survey variables. In the case of a longitudinal survey, these variables could be available in previous waves. Variables from (a), (b), (c) and (d) are successfully used in the literature to adjust for non-response. For example, using data from a number of surveys, Kreuter et al (2010) found that the inclusion of these variables in response propensity models that were used to derive non-response weights reduces the mean-square error (MSE) in measures of central tendency adjusted by the resultant weights. However, they found that very few of these variables are associated with the response propensity. In contrast, using Receiver Operating Characteristic (ROC) curves, Plewis (2011) assessed the impact of including these variables in the response propensity models. He found that their inclusion may improve the accuracy of the models (i.e. they are associated with response propensity), but they have little effect when adjusting for non-response. Also, Lynn (1996) showed how, in the Scottish School Leavers Survey (SSLS), information about the level of qualification gained at school which was available in the sampling frame was used to analyze the response rate in connection to making weighting adjustment. Lynn (1996) demonstrated the way in which the post code in the Health Survey for England (1994) was used to identify the area where the respondent lived as large urban/city centre, other urban/suburban or rural and then analyzed the response rate accordingly. In longitudinal surveys, variables from categories (a) to (d) are usually used to adjust for non-response in the first wave. After the first wave, it is common to use key variables from previous waves (i.e. category (e) variables) to analyze and/or adjust for non-response in later waves. Most research has found variables such as gender, race, age, socioeconomic status, income and level of education to be good predictors of the response propensity and hence powerful weighting variables. For example, Watson (2004) states that from wave 2 onwards in HILDA, variables such as gender, age, marital status, labour force status, health condition in a current wave are used to create non-response weights in the next wave. Similarly, age, gender, race, employment status, income and education are used in the BHPS weighting after wave 1 . Also, Siddiqui et al (1996) used proportional hazard regression in analyzing the factors influencing dropout in longitudinal school-based smoking prevention studies; race, tobacco knowledge and academic performance were found to be significant factors. Kroh (2009) indicates that, in GSOEP, characteristics measured in 2007 (wave 23) such as gender, age, job status, income and savings were used to predict the probability of re-interviewing in 2008. Both Becketti et al (1988) and Fitzgerald et al (1998) showed that, excluding young respondents, attrition is positively associated with old age. Investigating attrition in the BHPS, Uhrig (2008) found that housing tenure, marital status, size of household, gender, race, region, mode of interview, employment, number of children in household, financial situation, education, health, income and social isolation are all associated with attrition. In this thesis, the weights creation, is restricted to weighting variables from category (e), and a model-based method is used to create the standard and weights. The response propensity models in a given wave 2 include variables observed in previous wave 2. This enables taking into account changes in respondents characteristics which is likely to be reflected in the response propensities and hence in the weights. These cooperation variables can have great predictive power for panel attrition, because a sampled individual being hard to reach in the first wave interview can be considered as a negative reaction to the request to participate in the survey, thus increasing the probability of attrition in the subsequent waves. With a large number of candidate auxiliary variables, a desirable weighting adjustment method should be able to incorporate a large number of auxiliary variables without creating weighting adjustments that are too noisy to be useful.

3.4.6 Weighting in longitudinal surveys

A common practice in longitudinal surveys is that survey organizations prepare weights and include them in public use data files for use by analysts. Most of the household panel surveys implement a similar approach in terms of non-response weighting. To give an insight into this, in this section we describe and discuss this approach on the basis of the British Household Panel Survey (BHPS) since its data is used in this thesis.

The BHPS

Full details on the BHPS including the sample, survey instruments, fieldwork, measures and weighting procedures are well documented in Sadig (2011).

The BHPS was conducted in the period 1991 to 2008. It followed its sample members every year to conduct interviews. Its main purpose was to explore the dynamics of change experienced by the population in the UK. In addition, the BHPS was conducted so that secondary data users have micro-data sets available. These data sets can then be used to carry out a wide range of research across a range of social science disciplines, and for policy research. In general, the BHPS provides data in 9 main areas: labour markets, income, savings and wealth, household and family organization, housing, consumption, health, social and political values, education and training.

Eligibility to the BHPS was restricted to individuals who were residents in private households in the UK. Those who were not alive, not resident in the UK, or were in the UK but institutionalised (i.e. living in nursing homes, military bases or prison) were not eligible for the survey. Using the small user Postcode Address File (PAF) as a sampling frame, 8,217 addresses were drawn as original sample units. The frame included all countries in Great Britain except Northern Ireland. There were three stages of selection: using a systematic sampling technique, the first stage selected 250 postcode sectors from stratified listing of all sectors on the PAF as Primary Sampling Units (PSUs); in the second stage, fieldwork delivery points (equivalent to addresses) were selected from the resultant PSU from the previous stage using analogous systematic procedure; and a final selection stage was conducted by interviewers at the address level. During the selection of households, interviewers excluded non-residential addresses and institutions. A household in the BHPS was defined as “one person living alone or a group of people who either share living accommodation or share one meal a day and who have the address as their only or main residence”.

The first wave was conducted in 1991. Interviews were attempted with all household members who were aged 16 or over. This resulted in 10,248 individual interviews at wave 1. Subsequent to wave 1, the BHPS attempted following all sample members in wave 1 responding households and interviewing them as well as all new household members living with wave 1 sample members. Letters were sent to sample members, in subsequent waves, notifying them that the interviewer will call them within a week. Only adamant refusals were excluded from the fieldwork. Non-contacts were coded as such after six call attempts.

Weighting in the BHPS

The weighting in the BHPS is documented in volume A of the user manual (by Taylor, 2006).

To adjust for non-response, the BHPS calculates weights both at the individual and household levels. Our discussion will be limited to weights at the individual level, since the analyses here, and in most research, are done at the individual level. There are two types of weights: cross-sectional and longitudinal. Cross-sectional weights are available in every wave, but they are only suitable for cross-sectional analyses (single-wave analysis) in the corresponding waves. Longitudinal weights are available in every wave from wave 2 onwards. Longitudinal analyses that use data from a number of waves (any wave-combination of more than 1 wave) should use longitudinal weights from the last wave in the wave-combination in question. For example, to analyze data from wave 1, 10 and 18, or data from all waves up to wave 18, both scenarios should use the longitudinal weights at wave 18. In this section we describe the calculation of longitudinal weights.

At wave 1, there were two general types of weights: design weights and non-response weights. The design weights were derived to account for the different probabilities of selection due to the different stages of selecting the sample. These were calculated as the inverse of the probability of selection for every sample unit. However, our focus shall be on the creation of non-response weights. First, these were calculated at the household level using weighting classes. The variables used to identify the classes were region, socio-economic group (at the address level) and type of accommodation. In every class, the responding households were weighted by a factor that made their total number equal to the total number of responding and non-responding households in the class. A small number of cases within the responding households failed to respond at wave 1. However, information about these individuals were recorded during the household interview. To adjust for this, individual (within responding households) non-response weights were derived. A model-based method was used. By defining two outcomes: individual interview obtained =1; and individual interview was not obtained =0, a logit model was fitted. The variables used were age, gender, region, housing tenure, household size, marital status and employment status. For all responding cases, the weights were then defined as the inverse of the predicted probabilities from the model. These weights were then multiplied by the household non-response weight, and the resultant weights represent the individual non-response weights at wave 1. Note that BHPS does not release the design weights separately. The design weights were combined with the individual non-response weights from wave 1. Thus, the final analysis weight for a responding case at wave 1 is a product of the design weight and the individual non-response weight at wave 1 for that case. Final analysis weights in wave 1 represent the set of weights that is included in the BHPS wave 1 data file for analyses at the individual level.

In every subsequent wave, response was defined as responding in all waves up to and including the latest wave. In other words, both attrition and wave non-response were classified as absolute non-response. Non-respondents of unknown eligibility were treated as eligible non-respondents. The weights were then derived, every wave, (only) for those who responded at all waves up to the latest. Thus, the longitudinal weights at any wave are the product of subsequent weights accounting for losses between each adjacent pair of waves up to that point. Weighting in waves subsequent to wave 1, was done using weighting classes. A number of variables that were thought to be informative of non-response and of interest in the substantive analyses of BHPS data were used to form the classes. These variables include age, gender, race, employment status, income, education, region and tenure. At every wave, the method used variables from the previous wave. To make the process manageable, an automatic interaction detection program (SPSS CHAID) was used to create the weighting classes. The weight for respondents in a given class was defined as the inverse of the response rate of that class.

Most longitudinal surveys in the world apply similar approach with respect to weighting. The difference is that BHPS uses weighting classes to predict the probability of response, whereas other surveys (such as HILDA in Australia) apply a model-based method. In any case, this approach, which is typical in household panel surveys, will be referred to, throughout this thesis, as the standard weighting approach (SWA).

Apart from offering a single set of longitudinal weights at every wave, the SWA has the following principles:

(a) After the first wave, the response probability is estimated using a mixture of common variables (from the previous wave) on which all sample members have measurements, and all sample members are used as one set in this estimation.

(b) Response is identified as responding in all waves up to the latest, and therefore weights are only provided for units responding in all waves up to the last one. In other words, those who skip responding in at least one wave are also identified as non-respondents and, thus, do not have weights (i.e. $weight_j$=0; where j denotes attires, wave non-responders and complete non-participants).

However, the complexity of household longitudinal surveys raises concerns (linked to the above points a and b) with respect to using the SWA. In this thesis, we view these issues as limitations in the SWA. In particular, we investigate point ‘a’ above as a limitation in the SWA. In return, we design, discuss and evaluate a different alternative weighting approach, corresponding to this issue in the next two chapters of this thesis.

3.5 Limitations of the SWA: the research problem

Often, longitudinal surveys target large populations (typically a living population in a country). Sampled units from such populations are not usually homogeneous with respect to survey participation. Some sub-groups are more cooperative than others i.e. the response propensity may be driven by different factors for different sub-groups. Thus, non-response predictors (weighting variables) in these sub-groups may differ from the general non-response predictors.

In general, some variables are believed to be better weighting variables than others. For example, age and employment status are known to be good weighting variables because of their strong relationship with the response probability and most substantive analyses’ outcome variables; while religious beliefs is not generally considered a good weighting variable since it does not have a clear direct relationship with the response probability. Thus, using age and employment status together with other good predictors in weights creation generally yields a good set of non-response weights.

However, the same weighting variables may not be powerful in predicting the response probability in some sub-groups in the same sample. For instance, age and employment status might not effectively predict the probability of response in the sub-group of women aged 80 or over (because in this sub-group variation in age is minimal and all respondents are likely to be retired). In fact, variables such as religious beliefs may be a better predictor in this case. The point here is that using a common set of variables from the previous wave to create a single set of weights in the current wave – as in the SWA - does not necessarily result in a set of weights that can tackle non-response successfully in all sub-groups in the sample. Some sub-groups could use an alternative set of weights created from another set of variables. Some of these subgroups are important and are frequently used for analysis. This may be an alternative approach of weighting, but it needs practical evaluation.

The research in this thesis sets out to investigate whether the SWA appropriately tackles non-response error in different types of estimates from different types of analyses with respect to the raised limitation. The study aims to explore an alternative weighting approach (AWA) to deal with the choice of weighting variables for different subgroups. Also, the study evaluates the new approach of weighting as opposed to the standard weighting approach.

Chapter 4

Tailored weighting for subgroups

4.1 Introduction

A common practice in panel studies is to the logistic regression model to predict the probability of response and obtain non-response weights. The model is usually estimated using weighting variables e.g age & gender as well as all cases in the selected sample for which data is available on the weighting variables. This is a feature of the SWA in longitudinal surveys described in the previous chapter. Because all sample members are used in the process of modeling the response propensity and deriving the weights, it may be necessary for the SWA to use ‘generic’ weighting variables. These are successful in predicting response for the sample as a whole and, also, may be correlated with some of the survey key variables. Hence, variables that only distinguish response from non-response at a sub-group level may not be used in the SWA if they do not prove important at the full sample level.

Using variables that are correlated with the survey target variables is important in order to produce a set of weights that is successful in reducing non-response bias (Särndal and Lundström, 2005; Little and Vartivarian, 2003, Kreuter and Olson, 2011). The extent to which the bias is reduced is however also based on a good specification of the model in terms of using variables that significantly explain the variation in the response propensity in all sub-groups in the sample. If this is not observed the weights will not reduce non-response bias in estimates related to sub-groups where variation in response propensity either is not, or is poorly, accounted for by the weighting variables. In addition, the weights will reduce non-response bias to the maximum possible extent if they are used to adjust an estimate that is constructed using the set of respondents used to create the weights.

It is unlikely in practice for the SWA to account for the variation in the response propensity in all sub-groups in the sample, if it is based on just one weighting model, using all sample members and common weighting variables. This is so since, even in the same survey sample, the phenomena that cause non-response can differ across different sub-groups, both in terms of scale and type.

As an example, consider a survey that collects data from individuals belonging to different social classes. For the subgroup formed of teachers it is likely that the non-response rate will be lower than that in other sub-groups belonging to other social classes in the sample. This is due to the fact that individuals within academia may feel obligated to cooperate with the survey as an academic duty. It is however likely that the factors responsible for non-response in the sub-group of teachers and lecturers are rather different than the usual non-response predictors e.g gender, which could be more responsible for non-response in other sub-groups in the sample.

If a researcher wishes to obtain an estimate using only the subgroup of teachers & non-response weights, a model that is correctly specified to predict response probability in general i.e. SWA which is based upon all sample units, using variables that may be strongly correlated with the response propensity in many sub-groups in the sample but weakly correlated with the response propensity in the sub-group of teachers and lecturers, might result in a set of weights that successfully reduces non-response bias in many survey estimates but not necessarily reduce it in estimates which are constructed using only the set of teachers .

For any estimates that are restricted to the subgroup of teachers weights would be more effective if the model is estimated such that it specifically account for the variance in the response probability in the sub-group of teachers by using variables that are strongly correlated with the response propensity regardless of whether or not they also correlated with the response propensity in other sub-groups in the sample, and if it is estimated using only the set of teachers.

And because SWA is unlikely to predict response in all sub-groups in the sample, an alternative weighting approach is needed that sets modeling strategy that is able to account for the variation in the response propensity in a selected set of sub-groups. In this approach a number of different weighting models are estimated with the aim of explaining a larger proportion of variance in response propensity in certain sub-groups in the sample and use a particular set of variables which account for variation in the response probability in these sub-groups; and then estimate the model by using sample members from the sub-groups in question only. In this way the weights derived from each weighting model are expected to be more effective in dealing with non-response bias in their relevant sub-groups compared to weights derived from the SWA. Further more, if the sub-groups selected for this type of weighting represent some of the major domains in the sample, the resultant weights may also reduce non-response bias in estimates constructed from whole sample if they are put together appropriately. This strategy of weighting is will be denoted as the AWA in this chapter.

The objective of this research is to explore whether there is evidence that designing weights for specific sub-groups in the sample can significantly affect survey-based estimates from these sub-groups to the extent that they prove superior from the estimates produced through the SWA. The suggested approach AWA will be referred to as 'subgroup-tailored weighting approach' (S-TWA) and weights produced from this approach will be called 'tailored weights' (*TWs*).

The BHPS sample will be used to study the differences between the SWA and the proposed S-TWA. The investigation is based on creating weights using the SWA (*SWs*) and weights based on the AWA (*TWs*), and then compare estimates resulting from a substantive analysis that uses the *SWs* and *TWs.*

The S-TWA will be applied by selecting two sub-groups, from the BHPS sample, on which substantive analyses are intended to be done. The tailored weights will then be designed for these sub-groups by using variables that are thought to be correlated with the response propensity in the sub-groups under investigation regardless of whether or not these variables are also used in the SWA. In other words, that the S-TWA adds new variables to the common variables that are usually used in the SWA to create the *TWs.* The justification of using the new variables is that they are important predictors of the response in the selected sub-groups even if they are not of value in prediction of response in the other sub-groups. On the other hand, some of the variables used in the SWA may not be used in the S-TWA if they do not distinguish response from non-response in the sub-groups in question. The tailored weights can also be created by restricting the weighting models to the sets of sample members in the selected sub-groups.

4.2 The choice of sub-groups

The data used in this research was taken from the first eight waves of the BHPS[1] and cover the period 1991 to 1998. The investigation in was restricted to sample members aged 16 and over at the time who responded at wave 1. Furthermore the tailored weights were designed to target non-response bias in estimates related to two sub-groups of sample members. The first group consists of those who retired in 1991 or earlier. The second is composed of those born in 1965 or after.

Sample members in the first sub-group started the survey as retired individuals and hence not include respondents who retired at a later wave. The second sub-group contains sample members who were within the age group 16 to 26 at the start of the survey.

And although other types of sub-groups in the BHPS sample are also important, the selected sub-groups represent major domains in the sample. Furthermore, both of these sub-groups contain enough sample members to allow valid statistical investigation of the topic of this chapter. Also, the two sub-groups, include a set of sample members that is balanced in terms of gender, age (young and relatively old respondents) and labour market status (out of the labour force and working age individuals). In addition, a large number of extensive analyses may be conducted on the selected sub-groups. It is thus, worth exploring whether a set of weights tailored to these sub-groups differ from that produced by SWA.

The choice of the sub-groups divides the sample into three non-overlapping sub-groups namely retired sample members, sample members who were born in 1965 or after; and

[1] Some of the variables used in the analysis are not available across all waves.

non-retired sample members who were born before 1965, The S-TWA however concert rates on retired respondents and those who were born in 1965 or after. The tailored weights will be created to adjust for the longitudinal non-response up to wave 8. As a result, the weights will be appropriate for analyses, on the selected sub-groups, that use a balanced panel from the first eight waves of the BHPS. It is also worth noting that two of the major sub-groups in the sample, are used the weights are also expected to reduce non-response error in estimates related to the full sample analyses.

In section (4.3) we briefly outlines the construction of the *SWs.*

4.3 Weights from the SWA (*SWs*)

The process of constructing weights from the SWA (SWS) can be outlined as fallows.

Eight waves are available for the analysis. The *SWs* were created to adjust for the longitudinal non-response at wave 8. The response propensity is modeled at each wave conditional on responding at all of the previous waves. Those who are known to be ineligible by wave 8 were not included in these models, while those whose eligibility is unknown by wave 8 were assumed eligible cases and were included in the weighting models. This means that, the analysis was restricted to sample members known (or assumed) to be part of the target longitudinal population at wave 8 which the weighting here aims to represent. At each wave the model used variables from the previous wave. The variables used to model the response propensity are the usual weighting variables in the SWA, namely race, gender, age-squared, age, presence of children in the household, tenure, education, type of household, employment status, type of house, number in full-time employment in household and region.

Modeling of the response propensity started from wave 2 as the BHPS offers wave 1 non-response weights combined with the design weights. Equation (4.1) below explains the process formally.

$$Logit\ \Pr(R_{i,t}=1/\ C_{i,t-1}=1)=f\ (\textstyle\sum_j \beta_j Z_{ji}+\sum_k \beta_k X_{ki,t-1}) \quad (4.1)$$

Where t is the wave number for which the model is estimated (t=2, 3,…, T=8); i= 1, 2, …, $n_{1,\dots,t-1}$, where $n_{1,\dots,t-1}$ is the number of respondents who responded at every wave from 1 to $t-1$ and who are known or assumed as eligible by the time of wave 8; $R_{i,t}$ is the response status at time (wave) t for respondent i ($R_{i,t}$=1 if response is observed at wave t; $R_{i,t}$=0 if response is not observed at wave t); $C_{i,t-1}$=1 if $R_{i,b}$=1 for all values of b from 1 to $t-1$ (i.e. $C_{i,t-1}$=1 indicates that the model in wave t is conditioned on response in all of the previous waves); Z_{ji} is the set of time invariant variables for respondent i; $X_{ki,t-1}$ is the set of time variant variables for respondent i which are measured in wave $t-1$.

Table (4.1) displays the results of the final models of the SWA.

Table 4.1 Response propensity models based on the SWA (wave 2 to 8): modeling response in wave *t* conditional on responding in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Female	1.31*	1.22**	1.03	1.36*	1.13	1.29**	1.41**
White	1.69*	1.80**	1.23	1.71*	1.05	1.66*	1.83**
Age	1.07***	1.08***	1.10**	1.08**	1.11**	1.09**	1.12***
Age-squared	0.99***	0.99***	0.99**	0.99***	0.99**	0.99**	0.99**
House owner	1.43*	1.39**	1.28*	1.08	1.03	1.04	1.58**
Has GCE certificate or above	1.21*	1.04	1.13*	1.10	0.92	1.26*	1.19**
Employed	0.89*	0.79	1.16*	1.24*	0.87	1.39	1.31
Others people present during interview	0.87	1.28*	1.07	1.22	1.37*	1.19	0.94
Single-person HH	0.72	1.18	0.76*	0.94	0.88*	0.91	1.08
Household has children	1.39*	1.44*	0.89	1.06	1.58*	1.08	1.43
Lives in a flat	0.98	0.90	0.90	0.88	0.91*	0.87*	1.10
Lives in other type of house	1.12	0.91	0.75**	0.93	0.64*	1.21	0.87
1 or 2 persons are employed in HH	1.11	0.79*	0.94	0.89	1.22	1.31	1.07
3 + persons are employed in HH	0.93	0.63*	0.89	0.61*	1.12	1.16	0.88*
South-East	0.93	1.27	1.86*	0.84	1.33	1.45*	1.42
South-West	0.96	0.95	1.24	1.44	0.92	1.08	1.25
East Anglia	1.03	0.88	2.03*	1.86*	0.96	1.14	1.28
The Midlands	0.86	1.59	1.76	1.07	0.89	0.93	1.10
The North	1.23**	0.72*	1.51*	1.22	0.87	1.48*	1.36
Wales	1.44*	0.88	1.30	0.84	0.63*	1.26	0.79
Scotland	0.91	1,33	0.85	1.72*	0.49*	1.24	0.61*
n	9,593	8,699	8,218	7,863	7,496	7,152	6,878
Pseudo R^2	0.031	0.032	0.038	0.035	0.034	0.33	0.035

*Response is modeled conditional on responding in previous waves. The reference categories of the categorical independent variables are male, other non-white, not a house owner, does not have a GCE certificate or above, unemployed, others not present during interview, multi-person HH, household with no children, lives in a house, no one is in employment in household and London * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

As for the set of responding sample members in the 8 waves, the longitudinal *SWs* at wave 8 were calculated as the product of the inverse predicted probabilities from the models in table (4.2), and wave 1 non-response/design weights (provided by BHPS) as shown in equation (4.2).

$$SW_i = D_i * \prod_{t=2}^{8} r_{ti}^{-1} \qquad (4.2)$$

Where SW_i is the standard longitudinal weight at wave 8 for respondent *i*; r_{ti} is the predicted probability for respondent *i* from wave *t* model (*t*= 2, 3,…, 8);); *i*= 1,..., $n_{1,\dots,8}$ (where $n_{1,\dots,8}$ is the number of sample members who responded at every wave from 1 to 8); and D_i is wave 1 non-response/design weight for respondent *i*.

The distribution of the *SWs* is presented and discussed later on with the *TWs*.

4.4 Weighting variables for the subgroup-tailored weighting

Apart from the variables that are used in the SWA, for each of the selected sub-groups, some variables may be of particular interest in terms of predicting response in the sub-groups under study. These variables are not used in the SWA since they do not correlate with the response propensity for the sample as a whole. In this section, we propose and discuss two sets of the variables that will be used to create the tailored weights for our two sub-groups. The next section will embody explanation of the methodology that will be implemented in the S-TWA.

4.4.1 Proposed weighting variables for r*etired sample members*

The first proposed weighting variable is religion, since having a religion is a form of social participation. And although some authors suggest that social participation can negatively affect the contact attempt –by affecting the at-home pattern (e.g Lepkowski and Couper, 2002), others support the idea that social participation is an indication of higher human interaction levels so that a person who is socially interactive is more likely to cooperate and provide data for the survey (Groves and Couper, 1998). As for the BHPS sample, Uhrig (2008) noted that those who have religious beliefs are much more likely to respond than those who do not. He, however, found that this significant effect disappeares once other variables such as organizational participation are included in the model. The reason for this is that organizational participation is also an indicator of higher human interaction levels and hence indicator of survey cooperation. Some of the organizational participations however are more common among working-age respondents than retirees especially if they require a high physical activities and/or someone within the labour force. We assume in this research, that organizational participation such as joining sport clubs and professional organizations is more common amongst working-age respondents than their retired counterparts and hence it can only affect the estimated association between religion beliefs and survey participation of working-age respondents. As for the retired respondents, religion can then be considered as a good predictor of non-response.

The second variable considered is the respondent's energy compared to average at their age. We note here that, those who are more energetic than average at their age are expected to be more mobile and hence less likely to stay at home than those who have less energy. As a result in surveys that require making contact with respondents at their homes, it is more likely that less energetic people will be found at home than those with more energy. On the other hand, less energy than average may be associated with bad health leading to a lower level of cooperation or even refusal. Early works on non-response suggested that refusal for health reasons is common amongst elderly respondents (e.g Uhrig, 2008). For the sub-group of retired respondents (relatively old sample members), energy compared to average at the same age can be seen as an important indicator for both at-home pattern and health condition. Thus, whether or not this variable affects response propensity in the sub-group of retired respondents is worth exploring.

The third variable considered is whether respondent supports a political party**.** Since it is not clear whether there is direct association between political views and non-response prediction, few researchers used this variable. However, some authors (e.g. Groves and Couper, 1998) believe that those who have political views, may be more aware of the government's role in the society and therefore may feel more obligated to provide data for the survey. Some researches on political engagement suggests that it is lower amongst persons in working-age. This may be due to the fact that working-age respondents often do not have the time to engage with politics (Brandon, 2012). Retines,on the other hand, do not often have time problems; and do have the time to participate in politics. Actually they may feel the need to be socially interactive and therefore may participate in politics. Moreover, retirees could support and vote for a political party for reasons such as protecting the valuable benefits they receive from the government. As a result the assumption that supporting a political party can influence response propensity and it is more frequent amongst retired respondents in made (Brandon, 2012), and this variable was considered as a good weighting variable for retired respondents.

The fourth variable used is the subjective financial situation, studies on non-response has shown that there exists a positive relationship between financial position and response propensity (Groves and Couper, 1998; Fitzgerald *et al.*, 1998; Lepkowski and Couper, 2002). Those who are in better financial positions are more likely to respond than those who are less well off. For the BHPS sample, however the evidence for subjective financial situation is in contradiction with the general financial findings. Previous research on subjective financial position on the BHPS has confirmed that those who subjectively report themselves as being in better financial positions are less likely to respond than those who report themselves as being in worse financial positions (Uhrig, 2008). Nonetheless, the effect of subjective financial situation might change and confirm evidence from the general non-response literature once some sub-groups in the sample are controlled for (i.e. when the investigation is only done on retirees for example). For this reason, subjective financial situation was added to the set of weighting variables of retired respondents.

The final variable is having access to a car. Having access to a car for personal use is considered an indication of wealth and a good financial situation (Uhrig, 2008), and for retired respondents, it is looked at as an indicator of a good health. As a result, this variable was added here under the assumption that it is indicative of good health status and good financial situation.

4.4.2 Proposed weighting variables for *those who were born in 1965 or after*

Five weighting variables were used for those who were born in 1965 & over.

The first is liking the neighbourhood. As it indicates attachment to current neighbourhood. The feelings of respondents about their settlement in a neighbourhood are indicative of whether they will continue to live in that neighbourhood, and hence of the likelihood of locating and contacting them successfully. Studies have shown that younger respondents are more likely to move house (Uhrig, 2008). This variable is therefore likely to have a distinctive effect on the response propensity for those who were born in 1965 or after compared to their counterparts' sub-groups. Thus, this variable was added to the weighting variables of this sub-group.

The second variable is school leaving age. In the United Kingdom (UK) most people leave school at the age of 15 or 16. However, there are some exceptions. This may occur, for example, due to coming to the UK at the age of six and having to start school a year or two later than the average starting age (five years old). Circumstances in which one has to leave school at a different age than the average person may affect one's tendency to participate in the survey. Regardless of the nature of these circumstances, their existence can be expressed through the school leaving age. In this research, it is assumed that the effect of the circumstances associated with the school leaving age on survey participation fades over time. In other words, the effect is stronger at a younger age than at an older age. This is because living longer enables one to experience more life-events that may reduce any influence on survey cooperation due to the reasons why they left school at a different age than the average person. Thus, the relationship between school leaving age and non-response maybe of more interest for those who were born in 1965 or after than for those who were born before 1965.

Having children is the third variable proposed. Published works on non-response suggests that the presence of children in the household is positively associated with survey response (Groves and Couper, 1998; Lepkowski and Couper, 2002; Uhrig, 2008). This is so irrespective of whether these children are the respondent's own children. The reason is that, households with children are more settled and less likely to move house, and even if they do, they are easier to relocate and contact. This is especially important for younger respondents who are more mobile and less settled. Therefore, an item that measures if the respondent has their own children within the household for those who were born in 1965 or after can be considered a good weighting variable for this sub-group. This is because of its distinctive effect on the response process of those who were born in 1965 or after.

Subjective financial situation is the fourth variable proposed. As was mentioned earlier the evidence for subjective financial situation in the BHPS is in contradiction with the general financial findings (in the BHPS those in better financial positions are less likely to respond than those in worse financial positions). It is thus worth testing the effect of subjective financial situation on the response propensity of those who were born in 1965 or after too.

The fifth & last variable is having access to car**:** Aside from being indicative of wealth, having access to a car may have a distinctive effect on younger survey participants. It can be argued that having access to a car may affect the contactability of younger respondents. Therefore, this was included in the set of weighting variables of those who were born in 1965 or after.

4.5 The subgroup-tailored weighting approach (S-TWA)

In what follows we aim at incorporating the proposed weighting variables to construct a set of weights that is tailored to two sub-groups in the sample: retired sample members and sample members who were born in 1965 or after. There are at least two approaches to achieve this the interaction-based approach & modeling the non-response propensity separately for each sub-groups.

***Interaction-based approach*:**

In this approach the response propensity is modeled as done in the SWA, but interactions of the proposed variables for the tailored weighting for the two sub-groups under investigation are added to the models. For example, to capture the effect of religion (one of the proposed variable for the S-TWA) on the response propensity of retired sample members, one may add an interaction term of the variable that indicates whether a sample member is retired, and the variable that measures religion, to all of the weighting models estimated in the SWA.

When a response propensity model is estimated interaction effects may be tested. However, including interactions in the response propensity models that are used to derive non-response weights is not a common practice amongst survey researchers (Brick, 2013)[2]. Most survey organizations tend to rely on main effects when estimating their

[2] Unlike panel studies, some cohort studies such as the 1958 National Child Development Study (NCDS) in the UK used interactions to model the response propensities (Hawkes and Plewis, 2006).

response propensity models for construction of weights. Even in cases where interactions were used, the some studies suggest that weighting models with interaction effects have similar outcomes to weighting models with only main effects (e.g. Schouten, 2004). This may be due to the fact that, even when the interaction effects are used, they are only considered between the standard non-response variables (variables that affect response probability for most sample members). For variables that only predict non-response at a sub-group level, these might result in a different outcome.

Once all necessary interactions are included in the weighting models, the tailored weights can then be calculated, as usual, as the product of the inversed predicted probabilities from the estimated models. This approach has the advantage of being relatively straightforward to apply. Furthermore, it results in a single set of weights that is tailored to retired respondents and those who were born in 1965 or after. However, it may have some drawbacks. First, if there are many variables suggested for the S-TWA for each sub-group, the number of interaction terms becomes excessively large so that they can not all be included in one model. This is especially so if some of the proposed variables for the S-TWA are categorical variables with many categories (more than 2 categories). When too many interactions are included in the weighting model, this may result in less statistical power for other important variables in the model. Second, it uses all sample members to model the response propensity, including those who are not in the sub-groups under investigation. Thus, some variables, from the SWA, which are not associated with the response propensity in the sub-groups in question, will be kept in the weighting models because they may be associated with the response probability in other sub-groups in the sample, and hence, they will be used in the tailored weighting. Ideally

& as stated before, variables that do not distinguish response from non-response in the selected sub-groups should be excluded from the creation of the tailored weights for these sub-groups regardless of whether or not they predict response in other sub-groups. If such variables are kept in the model on the ground that they predict response in other sub-groups even if they are not important predictors for the sub-groups under investigation, the resultant weights in this case may not be fully tailored to the sub-group in question, they are, to an extent, standard.

***Model the response propensity separately for each sub-group* approach:**

In this approach the response propensity is modeled separately for each sub-group in the sample i.e separate weighting models are estimated for each sub-group). The subgroup-specific weighting models will only use variables that are associated with the response probability in the relevant sub-group. As a result, the set of weighting variables for a given sub-group may not include variables from the SWA that do not predict response in the given sub-group, and include only the variables that are proposed for the S-TWA for the sub-group. The weights will then be calculated, separately for each sub-group, as the product of the inversed predicted probabilities from the sub-group estimated models. Applying this approach will thus results in a subset of tailored weights for each sub-group. Which can the be combined to form an overall set of *TWs.*

It is possible that the two approaches yield similar results. The second approach, my however be superior especially if many categorical variables are suggested for the S-TWA since it will be more practical in this case, and also because it allows exclusion of the variables that are not significant at the sub-group level.

In this study we apply both methods of subgroup-tailored weighting as AWAs. This strategy enables us to see whether the two approaches can result in different outcomes. We refer to the tailored weights resulting from the *interaction-based approach* as TWs_1, whereas weights resulting from *modeling the response propensity separately for each sub-group* are denoted as TWs_2. The construction of the TWs_1 and TWs_2 is explained in the next section.

4.6 Construction of the tailored weights (*TWs*)

4.6.1 Interaction-based approach

Here two indicators are created one for retired sample members and the other for those who were born in 1965 or after (1=retired, 0=non-retired; and 1=born in 1965 or after, 0=was not born in 1965 or after). The same weighting models of the SWA is used and interactions of each indicator are added and its relevant proposed weighting variables introduced in section 3.4. Table (4.2) gives the results of modeling the response propensity using this approach. Note that we do not include 'age' and 'household with children' in these models as there are two variables used in the tailored weighting that can substitute for these ('born in 1965 or after and' 'has their own children' respectively).

The results regarding the variables proposed for the S-TWA here are similar to those from modeling the response propensity separately for each sub-group which will be presented next. Thus, the effect of including these variables in the weighting process will be discussed in detail in the next section. However, the major findings here will be highlighted.

With respect to the standard weighting variables (i.e. the variables used in the SWA), most variables have the same effect on the response propensity as in the SWA.

Second, none of the main effects of our new added variables appear to be significant in the models displayed in table (4.2) (with the exception of ‘has their own children’ as this substitutes for ‘household with children’). The significance of these variables is reflected in their interactions with the indicators of the two sub-groups in question. For example, the variables: religious and likes their current neighbourhood do not seem to be significant in predicting response for the sample as a whole. However, the interactions of these variables with retired sample members and those who were born in 1965 or after, respectively, appear to be significant suggesting that these variables are important in predicting response in the sub-groups under investigation.

This result confirms our clain that non-response process may be different in the selected sub-groups than in the sample in general. Furthermore, it shows that some of the factors responsible for non-response in these sub-groups are different than the factors responsible for non-response in the other subgroups in the sample. In addition and based on this finding, one can expect our proposed variables to be significant when modeling the response propensity separately for each sub-group as will be shown next.

Table 4.2 Response propensity models based on the AWA (interaction-based): modeling response in wave *t* conditional on responding in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Female	1.21*	1.28***	0.93	1.31**	1.12	1.27*	1.35**
White	1.46*	1.53***	1.30*	1.42*	0.88	1.51*	1.49**
House owner	1.15*	1.19*	1.22*	1.31**	0.89	1.33**	1.13
Has GCE certificate or above	1.22*	0.92	0.87	1.08	1.26*	1.11	1.39**
Employed	0.85	1.09	1.06	1.23*	0.91	0.88	1.19*
Other people present during interview	1.05	0.87	1.08*	1.04	1.39**	1.33*	1.07
Single-person household	0.93*	1.30	0.88**	1.10	0.96	0.76*	0.92
Lives in a flat	0.96	1.14	0.89	0.83*	1.03	0.87*	0.92
Lives in other type of house	0.88	0.81	0.69*	0.92	0.94	0.72*	1.05
1 or 2 persons are employed in HH	1.05	0.93	0.92	0.89*	1.08	1.11	1.20
3 + persons are employed in HH	0.91	0.68*	1.11	0.88	0.92	0.65*	0.93
South-East	0.94	0.93	1.36*	1.40*	0.77	1.07	0.85
South-West	1.29	0.95	1.19	1.17	1.20	0.88	1.05
East Anglia	0.91	0.93	1.66*	1.51*	0.89	1.03	1.33
The Midlands	0.97	0.89	1.22	1.26	0.86	1.11	0.91
The North	1.03	0.87*	1.34*	1.22	0.81	1.41*	1.19
Wales	0.93	0.74*	1.58*	1.02	0.67*	0.95	0.86
Scotland	0.89	0.81	1.29	1.44*	0.31**	1.10	0.83*
Retired	1.21*	1.26	1.18*	0.92	0.53*	0.71*	1.06
Religious	0.84	1.05	1.11	1.15	0.87	1.01	1.19
Retired * religious	1.22*	1.15*	1.49	1.27*	1.68*	1.11	0.87
Has more energy than average at their age	1.13	1.09	0.87	1.11	1.18	0.86	1.09
Has less energy than average at their age	0.92	1.06	1.12	1.11	0.94	0.90	0.89
Retired * has more energy than average at their age	1.11	0.97	1.59*	1.07	0.88	1.68*	1.39*

Table 4.2 (continued)

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6
Retired * has less energy than average at their age	0.51***	0.39***	0.44**	0.38**	0.42**
Supports a political party	1.08	0.93	1.11	1.05	0.92
Retired * supports a political party	1.31*	1.30*	1.16	1.19	1.34*
Financially okay	0.89	0.81	0.78	0.91	0.85
Financially struggling	0.88	0.93	0.82	0.87	0.77
Retired * financially okay	0.96	0.94	0.87	1.12	0.90
Retired * financially struggling	1.10	0.85*	0.89*	1.24	0.84*
Has a car	1.29	1.31	1.13	0.94	1.03
Retired * has a car	1.26*	1.19*	0.90	0.94	1.11*
Was born in 1965 or after	0.81*	0.79*	1.03	1.16	1.36*
Likes the current neighbourhood	1.22	1.28	1.10	1.31	1.22
Born in 1965 or after X likes the current neighbourhood	1.14*	0.73	1.10*	0.77	1.26
Left school aged 14 or less	0.71	0.86	0.89	1.18	0.82
Left school aged 17 or over	1.27	1.22	1.30	1.05	0.89
Born in 1965 or after X left school aged 14 or less	1.15	0.81	0.73*	0.49*	0.91
Born in 1965 or after X left school aged 17 or over	1.17	1.19	0.84*	1.18	0.88*
Has their own children	0.92	1.25*	1.08	1.22*	0.97
Born in 1965 or after * has their own children	1.10	1.23*	1.07	1.26*	1.06
Born in 1965 or after * financially okay	0.92	0.89	1.02	1.12	1.11
Born in 1965 or after * financially struggling	1.15*	0.75	1.28*	0.93	1.34*
Born in 1965 or after * has a car	0.96	0.72*	0.83*	0.91	1.08
n	9,593	8,699	8,218	7,863	7,496
Pseudo R^2	0.030	0.033	0.036	0.035	0.032

For the set of responding sample members in the 8 waves, the TWs_1 at wave 8 were calculated as the product of the inversed predicted probabilities from the models in table (4.2), and wave 1 non-response/design weights as shown in equation (4.3).

$$TW_{1i} = D_i * \prod_{t=2}^{8} r_{ti}^{-1} \tag{4.3}$$

Where TW_{1i} is the *interaction-based* subgroup-tailored weight at wave 8 for respondent *i*; r_{ti} is the predicted probability for respondent *i* from wave *t* model (*t*= 2, 3,…, 8); *i*= 1,..., $n_{1,..,8}$ (where $n_{1,..,8}$ is the number of sample members who responded at every wave from 1 to 8); and D_i is wave 1 non-response/design weight for respondent *i*.

The distribution of the TWs_1 will be presented and discussed, together with the TWs_2 and the *SWs* in section 4.7.

4.6.2 Modeling the response propensity separately for each sub-group

To model the response propensity separately for each sub-group three different groups of weighting models were estimated. Remember that we have three sub-groups in the sample: retired sample members, sample members who were born in 1965 or after, and non-retired sample members who were born before 1965. However, the S-TWA focuses on the first two. Hence, for each of these two sub-groups, the weighting models excluded some of the variables used in the SWA which are not important in the given sub-group in terms of predicting response, and included the relevant proposed weighting variables. This makes the sets of weighting variables used in the creation of the TWs_2 for each of the selected sub-groups different from each other and from the set of variables used in the SWA. As for the weighting models of the third sub-group (non-retired who were born before 1965), this used the same variables from the SWA.

In the remaining part of this section we discuss, in detail, the results from modeling the response propensity separately for each sub-group. Our discussion here will be limited to the variables proposed for the SWA as the other variables result in similar results to the SWA.

Modeling response propensity for retired sample members

Some of the variables that were used in the SWA were excluded from this analysis, as described below and new variables were added. The added variables are our proposed variables for the S-TWA for the retired sample members. Furthermore, the weighting models were estimated using only the set of retired sample members. The results of the weighting models of the retired sample members are presented in table 4.3.

Excluded variables

The dropped variables are employment status and number in employment in household. Employment status is an important factor that predicts response propensity in the analysis of non-response because it is a good predictor of the probability of contact. Normally, those who are in full-time employment are more difficult to contact since they are less likely to be at home (Groves and Couper, 1998). In spite of that, employment status was excluded from the set of weighting variables in this case as all of the sample members in this sub-group are retired.

In any survey that contacts sample members at their home, a successful contact attempt with any household depends on whether some (or at least one) of the household members are (is) actually at home to respond to the contact attempt. So, the number of household members in employment can be negatively associated with successful contact attempts. Consequently, households with more individuals in full-time employment are less likely to respond compared to households that have less number of individuals in full-time employment. This is confirmed by our results from the SWA in table (4.1). However, dealing with retired sample members guarantees that there is at least one household member who is not in a full-time employment and hence it is more likely to successfully establish contact in this case. Since this applies to all retirees, this variable was excluded from the choice of weighting variables for retired sample members.

Added variables (proposed for the tailored weighting of retired respondents)

Five variables were added. ***The first variable added is religion.*** Religion was included in the model as a categorical variable with two categories. These are religious and non-religious (reference category). Most of the models in table (4.3) show that those who have a religion are more likely to respond than those who do not have a religion.

The second variable added is respondent's energy compared to average at their age.

This variable was included in the weighting models as a dummy variable with three categories, These are has the same energy as average at the same age (reference category), has more energy compared to the average at the same age and has less energy compared to the average at their age. Table 4.3, shows that most of the models indicate that those who have more energy than average are more likely to respond than those who have the same energy as average. In contrast, sample members with less amount of energy compared to the average at their age are less likely to respond than those who have the same energy as average at their age. This is due to the fact that ‘energy’ may be a strong indicator of the physical ability of a retired sample member to take part in the interview. Thus, retired individuals with more energy than average are likely to be in a good health condition, which may in turn increase the likelihood of successfully conducting the interview with older respondents. Those who have less energy compared to people at their age, it is less likely that they will be cooperative compared to those with same energy as average.

The third variable added is whether respondent supports a political party. This is a categorical variable with two categories: supports a political party and does not support a political party (reference category). Our response propensity models in table (4.3), show as expected, that when this variable is significant, those who support a political party are more likely to respond than those who do not. This is in line with our hypothesis suggesting that retired sample members who have political views may feel more obligated to respond to the survey.

The fourth variable is subjective financial status. The BHPS measures the subjective financial situation by asking respondents the question “how well would you say you yourself are managing financially these days?” In turn, respondents have to report their financial situation by selecting one of these options: living comfortably, doing alright, just about getting by, finding it quite difficult and finding it very difficult. Rearranging these options by combining the second option with the third, and the fourth option with the fifth, subjective financial situation was included in the models as a categorical variable with three categories: having a good financial situation (reference category), financially okay and financially struggling. The results suggest that, for retired respondents, those who are better off are more likely to respond than those who are less well off. The models indicate that both those who are financially okay and those who are financially struggling are less likely to respond than those with a good financial situation. These results are similar to the general findings of the effect of wealth on the response propensity. However, recall that the evidence from the BHPS (for the whole sample) regarding financial situation is in contradiction with this finding (Uhrig, 2008). Thus, confirming our hypothesis, the results here indicate that the effect of financial situation on the response propensity is different for retired respondents than for the rest of the sample.

The final variable added is having access to a car. This was included in the model as a categorical variable with two categories: has a car and has no car (reference category). Most of the models in table 4.3 show that retired respondents who have access to a car are more likely to maintain response than those who do not have access to a car. Our explanation for this is that, for retired sample members, having a car for personal use is indicative of a good physical health and relatively good financial situation.

Table (4.3) Response propensity models for retired respondents: modeling response in wave *t* conditional on responding in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Female	1.12**	1.19*	1.02	1.23*	1.40	1.38	1.19
White	1.31	1.68*	1.37*	1.13	1.26	1.64	1.59
Age	1.01*	0.99	1.25***	1.11***	1.18***	1.19***	1.25*
Age-squared	0.95*	0.98	0.99***	0.99**	0.99**	0.99***	0.99*
Home owner	1.30*	0.99	0.84	0.97	1.14*	1.47	0.84
Has GCE qualification or above	1.07	0.98	1.63*	1.20	1.41*	1.28	1.41
Others present in interview	0.88	1.18*	1.22*	1.14	1.41	1.13	1.78
Single-person household	0.99	0.97*	0.78**	1.17	0.64	0.73*	1.16
Household with children	1.32*	0.87	0.83	1.62*	0.73	0.79	1.11*
Living in a flat	0.60	0.73*	0.75	1.18	1.02	0.97	0.29
Living in other type of house	1.32	0.92*	0.86	0.70*	0.60	0.69*	0.66*
South-East	1.02	1.01	0.55	1.20	0.52	2.11	2.85
South-West	0.98	1.16*	1.12	1.39	0.63	1.27	1.93
East Anglia	1.41*	0.88	1.23*	1.35	0.67	1.51	1.20
The Midlands	1.32	0.73	1.01	1.43	0.46	1.77	1.82
The North	1.18	1.08	1.05	1.12	0.65	1.21	1.40
Wales	0.89	1.17	0.86	1.14	0.40*	1.42	1.24
Scotland	1.39	2.17	1.05	3.74*	0.50	0.96	0.87
Religious	1.03*	1.56*	1.34*	1.39*	1.16*	1.84	0.82
Has more energy compared to average at their age	0.92	1.06	1.24*	1.37	1.60	1.33*	1.46*
Has less energy compared to average at their age	0.46***	0.48***	0.49**	0.55**	0.53*	0.61	0.88
Supports a political party	1.08*	1.12*	0.86	0.94	1.10*	0.97	1.09*
Financially okay	0.90	1.10	1.13	1.38	1.39	0.89*	1.29
Financially struggling	1.04	0.87*	0.85*	0.93	0.88*	1.06	0.79*
Has a car	1.08*	1.16*	0.52	1.01	1.21*	0.65	1.20*
n	1,712	1,647	1,594	1,550	1,496	1,457	1,418
Pseudo R^2	0.037	0.038	0.039	0.035	0.038	0.036	0.036

* The entries are odds ratios. In every wave response is modelled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. The reference categories of the categorical independent variables in the table are male, non-white, not a home owner, does not have a GCE or above degree, others not present when interviewed, multi-person HH, household with no children, living in a house, London, non-religious, has the same energy as average as their age, does not support a political party, having good financial situation and does not have a car * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

For the set of responding retired sample members in the 8 waves, the tailored weights at wave 8 were calculated as the product of the inversed predicted probabilities from the models in table (4.3), and wave 1 non-response/design weights as shown in equation (4.4).

$$TW_{RRi} = D_i * \prod_{t=2}^{8} r_{ti}^{-1} \qquad (4.4)$$

Where TW_{RRi} is the tailored weight at wave 8 for retired respondent *i*, based on modeling the response propensity separately for retired sample members; r_{ti} is the predicted probability for retired respondent *i* from wave *t* model (*t*= 2, 3,…, 8); *i*= 1,..., $n_{1,\dots,8}$ (where $n_{1,\dots,8}$ is the number of retired sample members who responded at every wave from 1 to 8); and D_i is wave 1 non-response/design weight for retired respondent *i*.

Modeling response propensity for those who were born in 1965 or after

As in the case of modeling the response propensity for retired respondents, the weighting models for those who were born in 1965 or after were estimated by changing some of the weighting variables used in the SWA and by using the set of sample members who were born in 1965 or after. The results of the weighting models of those who were born in 1965 or after are displayed in table (4.4).

Two variables were dropped.

The first is age.

Age is an important factor in predicting non-response. Published research indicates that, in general, elderly people are more likely to refuse to participate in the survey than younger respondents (Groves and Couper, 1998; Lepkowski and Couper, 2002).

However, other research suggests that the youngest respondents in the sample are more difficult to locate as they have a higher tendency to move house, and even if they are located, they are still difficult to contact because they are less likely to be at home (Stoop, 2005). This pattern is very common among the vast majority of younger sample members. In this research, respondents who were born in 1965 or after fell into the age group 16-26 by the time the first wave of BHPS was conducted. This age group forms the youngest age group in the sample. But, preliminary analysis for this age group showed that age is not an important factor to predicting non-response within this age group. Hence, the weighting models for those who were born in 1965 or after were estimated without including the variable age.

The second variable dropped is whether children in household.

This variable was used to estimate the weighting models in the SWA. It shows whether if there are children within the household. This is regardless of whether these children are the respondent’s own children (i.e. could be nephews, nieces, etc…). Non-response theory suggests that the presence of children in the household is associated with high levels of response. This is because the presence of children in the household indicates more social integration (e.g. taking the kids to school or nursery) and hence it is easier to locate and contact households with children than single-person households or households with no children (Groves and Couper, 1998; Uhrig, 2008). However, one of the proposed weighting variables for those who were born in 1965 or measures the respondent’s own children in the household. This variable somewhat substitutes for the presence of children in the household and therefore the latter was excluded from the weighting model of those who were born in 1965 or after.

Five variables that are added for the tailored weighting of those born in 1965 & after:

The first variable added is liking the neighbourhood**:**

It was included in the models as a categorical variable with two categories, likes their current neighbourhood and does not like their current neighbourhood (reference category). As shown in table 4.4, when this variable is significant, it indicates that those who like living in their neighbourhood are more likely to respond than those who do not like living in their neighbourhood. This result indicates that one's attachment to the neighbourhood where they reside may be particularly important in predicting response for those who were born in 1965 or after. In general, individuals who are not attached to their residence neighbourhood are likely to move house and hence may be difficult to track and re-establish contact with. However, this is especially more likely for younger sample members (those who were born in 1965 or after in our case) who are usually more mobile compared to their older counterparts.

The second variable added is school leaving age**.**

In order to measure this variable, BHPS sample members were asked the following question: "how old were you when you left school". In return, if not still at school, respondents reported the age at which they left school. The reported ages range between 9 and 22. These answers were categorized into three categories: left school aged 14 or below, left school aged 15 or 16 (reference category) and left school aged 17 or above. At the time of wave 1, there was a small number of respondents who were still in school. This group of sample members does not allow valid estimation of the weighting models if they are treated as a separate category. This is especially the case in the weighting models after wave 2 as more cases from this category leave school as time goes on. Thus, these cases were classified with the category ‘left school aged 17 or above’ (since everyone in our sample aged 16+ at wave 1, eventually those who were still in school at the time of wave 1 will have left school aged 17+). Most of our models here suggest that both those who left school aged 14 or below and those who left school aged 17 or above are less likely to respond than those who left school aged 15 or 16.

***The third variable is having children in the househod*:**

In the BHPS data set there is a variable that refers to the number of the respondent’s own children in the household. The value of this variable ranges from between 0 and 9. This variable was used to indicate whether the respondent has children or not. It was categorized into two categories: has their own children in household (by combining the numbers from 1 to 9 in one category) and does not have their own children in household (reference category). As expected, the results suggest that those who have their own children within the household are more likely to respond than those who do not have children in the household.

***The fourth variable added is subjective financial situation*:**

As in modeling the response propensity for retired sample members, financial situation here was included in the models as a categorical variable with three categories: having a good financial situation (reference category), financially okay and financially struggling. Unlike the findings for retired sample members, the evidence here suggests that those who are less well off are more likely to respond than those who are better-off. This result confirms that financial situation is indeed an important factor for predicting response for both retired sample members and those who were born in 1965 or after. However, and more importantly, it shows that the effect of this variable is different for the two sub-groups. Thus, a weighting strategy like the SWA which might not recognize this as it estimates its weighting models by assuming that the effect of such variable is similar for all sub-groups may result in a set of weights that does not properly adjust for non-response in estimates of financial phenomena which are related to the sub-groups in question.

***The last variable added is having access to car*:**

It a car was included in the model as a categorical variable with two categories: has a car and has no car (reference category). The results for this variable indicate that those who have a car for personal transport are less likely to respond that those who do not have a car. One possible explanation for this is that, for younger sample members (those who were born in 1965 or after), having a car may be a factor that stimulates the 'not at home pattern'. Thus, young sample members who have a car may be less likely to be contacted successfully than those who do not have a car.

Table 4.4 Response propensity models for those who were born 1965 or after: modeling response in wave *t* conditional on responding in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Female	1.09	1.36*	1.66	1.33	1.20*	1.95	1.25*
White	1.27**	1.18*	1.22*	1.07	0.58	1.24	1.17*
Home owner	1.12*	0.95	1.21	1.17	1.71	1.39**	1.57
Has GCE degree or more	1.28	1.08*	1.49	1.19*	1.29	1.63	1.46
Employed	0.71	0.94	1.07*	1.15*	1.37	1.13*	1.17
Others present in interview	1.28	1.45	1.47	1.54*	0.83	1.14*	0.99
Single-person household	0.80*	1.07	1.21	0.98	0.63*	0.87	0.69*
Living in a flat	0.79	0.65	0.65*	0.59	1.32	0.72*	1.83
Living in other type of house	0.84	0.66	0.55*	1.53	1.39	0.90	1.29
1 or 2 persons in employment	1.20	1.01	1.11	1.10	0.64*	0.69	0.77*
3 + persons in employment	0.90	0.53*	1.27	0.99	0.43*	0.78	1.93
South-East	0.69	1.15	2.31	0.89	1.53	0.61	0.37
South-West	1.13*	1.45	2.00	1.08	3.23	0.42	0.81
East Anglia	1.74*	1.21	1.19	1.25	1.07	2.37*	1.46*
The Midlands	1.05	1.21	1.16	0.69	1.67	0.70	0.64
The North	1.38	1.02	2.53	0.91	2.37	0.80	0.69
Wales	0.70	0.83	2.22	0.48*	1.30	0.53	0.39
Scotland	1.20	1.12	1.55	0.62	1.06	0.54	0.32
Likes their current neighbourhood	1.20*	0.98	1.42*	0.69	0.89	1.56*	0.97
Left school aged 14 or less	0.91	0.56	0.64*	0.40*	0.51	0.53*	0.49
Left school aged 17 or above	0.89	0.84	0.62*	1.04	0.77*	0.76*	1.06
Has their own children	0.86	1.18*	1.52	1.55*	0.97	1.63*	1.50
Financially okay	0.89	0.97	0.93	1.31	1.48	1.22*	1.11*
Financially struggling	0.84	0.83	1.14*	0.88	1.21*	1.25	0.90
Has a car	1.15	0.69*	0.86*	1.04	1.05	0.79*	1.18
n	1,933	1,862	1,798	1,757	1,695	1,651	1,576
Pseudo R^2	0.030	0.034	0.035	0.038	0.036	0.039	0.035

*The entries are odds ratios. In every wave response is modeled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. The reference categories of the categorical independent variables in the table are male, non-white, not a home owner, does not have a GCE or above degree, unemployed, others not present when interviewed, multi-person HH, living in a house, no one in employment in HH, London, does not like their current neighbourhood, left school aged 15 or 16, does not have their own children, having good financial situation and does not have a car * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

For sample members who were born in 1965 or after and who responded in the 8 waves, the tailored weights at wave 8 were calculated as the product of the inversed predicted probabilities from the models in table (4.4), and wave 1 non-response/design weights as shown in equation (4.5).

$$TW_{1965i} = D_i * \prod_{t=2}^{8} r_{ti}^{-1} \qquad (4.5)$$

Where TW_{1965i} is the tailored weight at wave 8 for respondent *i* (who was born in 1965 or after) based on modeling the response propensity separately for sample members who were born in 1965 or after; r_{ti} is the predicted probability for respondent *i* from wave *t* model (*t*= 2, 3,…, 8); *i*= 1,..., $n_{1,..,8}$ (where $n_{1,..,8}$ is the number of those were born in 1965 or after who responded at every wave from 1 to 8); and D_i is wave 1 non-response/design weight for respondent *i*.

Modeling response propensity for those who are non-retired and born before 1965

The set of weighting variables used to estimate the weighting models for those who are non-retired and born before 1965 is the same as the set of weighting variables used in the SWA. However the models were only restricted to those who are non-retired and were born before 1965. Table 4.5 shows the results of modeling the response propensity in the 8 waves for this part of the sample. As expected, the results here are similar to the ones from the SWA. Overall, the results indicate that response is higher amongst females, white sample members, those with more education, employed individuals and members of multi-person households or households with children.

Table 4.5 Response propensity models for non-retired respondents who were born before 1965: modeling response in wave *t* conditional on responding in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Female	1.14	1.38**	1.14	1.19*	1.13	0.98	1.53*
White	1.29**	1.87**	1.44*	1.67*	0.70	1.83	1.33**
Age	1.03*	1.08***	1.07**	1.14***	1.10***	1.17***	1.24***
Age-squared	0.99*	0.99***	0.99***	0.99***	0.99***	0.99***	0.99***
Home owner	1.04	1.50**	1.38*	1.57**	1.13*	1.63*	1.21
Has GCE degree or more	1.25**	1.11	0.96	0.98	0.91	1.01	1.22*
Employed	1.07	1.02	1.25*	1.14	1.17*	1.08	1.20*
Others present in interview	1.08*	1.12	0.86	1.05	1.07	1.51*	1.49
Single-person household	1.29	0.91*	0.90	0.92	1.01	0.72*	0.83*
Household with children	1.25**	1.06	1.30	1.32	1.02	1.05	1.26*
Living in a flat	0.72	0.79*	1.09	0.87	0.77*	0.83	0.79
Living in other type of house	1.05	1.28	0.46*	0.89	1.49	0.65*	1.64
1 or 2 persons in employment	1.01	0.65*	0.93	0.96	1.25	1.08	0.71
3 + persons in employment	0.90	0.56*	0.81	0.61	0.58	0.34*	0.42*
South-East	1.09	0.80	1.37	1.75	1.41	1.19	0.80
South-West	1.18	0.78	1.12	1.32	1.58	0.91	0.88
East Anglia	0.92	0.84	2.05*	2.71*	0.89	1.34	1.17
The Midlands	0.92	0.72	1.25	1.42*	1.22	1.15	0.82
The North	1.31*	0.60	1.37	1.26	0.86	2.08**	0.89
Wales	0.90	0.56	1.36	1.34	1.70*	1.59	1.20
Scotland	0.84	0.45*	1.78	1.74	0.50	1.25	0.44*
n	5,948	5,190	4,826	4,556	4,305	4,044	3,884
Pseudo R^2	0.029	0.033	0.034	0.033	0.032	0.034	0.034

*The entries are odds ratios. In every wave response is modeled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. The reference categories of the categorical independent variables in the table are male, non-white, not a home owner, does not have a GCE or above degree, unemployed, others not present when interviewed, multi-person HH, household with no children, living in a house, no one is in employment in HH and London * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

For sample members who are non-retired and were born before 1965 (remaining sample), and who responded in the 8 waves, the tailored weights at wave 8 were calculated as the product of the inversed predicted probabilities from the models in table (4.5), and wave 1 non-response/design weights as shown in equation (4.6).

$$TW_{RSi} = D_i * \prod_{t=2}^{8} r_{ti}^{-1} \qquad (4.6)$$

Where TW_{2i} is the tailored weight at wave 8 for respondent *i* (who is non-retired and was born before 1965) based on modeling the response propensity separately for sample members who are non-retired and were born before 1965; r_{ti} is the predicted probability for respondent *i* from wave *t* model (*t*= 2, 3,…, 8); *i*= 1,..., $n_{1,..,8}$ (where $n_{1,..,8}$ is the number of those who are non-retired and born before 1965 who responded at every wave from 1 to 8); and D_i is wave 1 non-response/design weight for respondent *i*.

Since the three sub-groups in the analysis are non-overlapping, the sub-sets of tailored weights resulting from modeling the response propensity for each sub-group were then put together to form our second set of tailored weights (TWs_2) as follows:

$$TWs_2 = TW_{RR} \cup TW_{1965} \cup TW_{RS} \qquad (4.7)$$

The distribution of the SWs, TWs_1 and TWs_2

This section discusses and present the distribution of the weights resulting from the SWA (*SWs*) and the two methods of the S-TWA (TWs_1 and TWs_2). Table (4.6) presents the measures of central tendency and dispersion for the three sets of weights. For each set of weights, these statistics are presented separately for retired respondents, those who were born in 1965 or after and for all respondents.

In section of the standard errors of the three sets of weights under investigation, shows that these weights have very similar dispersion within all sets of respondents. This is confirmed by the coefficients of variations (CV) which are almost identical for the three sets of weights across the three groups of respondents. Thus, with the same amount of variation in all sets of weights, it seems reasonable to expect rather similar results in terms of precision for equivalent estimates constructed with *SWs*, TWs_1 or TWs_2.

However with respect to the average weight value, a different picture emerges. While TWs_1 and TWs_2 have the same average weight values across the three sets of respondents, *SWs* seem to have smaller weights sizes on average compared to the tailored sets of weights. This is also the case with the medians, and the first and third quintiles values indicating that, for most cases in the sample, TWs_1 and TWs_2 contains fairly larger weights compared to the *SWs*. These results suggest that the S-TWA resulted in somewhat different weights than the SWA in terms of the average weights value. Accordingly, we expect this to affect the magnitude of some of the estimates resulting from the S-TWA, possibly to an extent that makes them significantly different than their equivalent estimates resulting from the SWA.

Furthermore, we do not expect to find considerable differences between estimates resulting from TWs_1 and TWs_2 as the distributions of these two sets are very similar both in terms of dispersion and average weights value. This, in turn, suggests that our two approaches of sub-group tailored weighting may have similar effects on the resultant weights.

Table 4.6 The distribution of standard weights and tailored weights.

	Retired respondents			Born 1965 or after			The whole sample		
	SWs	TWs_1	TWs_2	SWs	TWs_1	TWs_2	SWs	TWs_1	TWs_2
Std.dev	0.66	0.68	0.69	0.67	0.68	0.69	0.63	0.65	0.66
Mean	1.98	2.12	2.10	1.79	1.86	1.87	1.58	1.65	1.64
CV	0.33	0.32	0.33	0.37	0.37	0.37	0.40	0.39	0.40
Min	0.49	0.54	0.59	0.40	0.46	0.44	0.32	0.35	0.33
Q1	1.61	1.71	1.73	1.42	1.58	1.55	1.26	1.31	1.36
Median	1.80	1.89	1.87	1.69	1.76	1.77	1.49	1.54	1.55
Q3	2.11	2.35	2.30	2.06	2.12	2.15	1.86	1.91	1.93
Max	6.37	10.86	10.05	6.06	7.24	7.75	6.88	11.16	11.89

*CV is the coefficient of variation (CV=Std.dev/Mean)

Chapter 5

Analysis and results

5.1 Introduction

In this part, we will be analysing the three sets of weight (TWs_1, TWs_2 and *SWs*). All these sets have been devised to fine-tune the collective non response among waves 1 and 8. This reflects that each member of the sample will have access to the weights of all these sets who will be responding to surveys first 8 waves. Hence, for this fundamental analysis, we will be using balance board of members who will be responding to waves 1 to 8 of the survey (i.e. 6,753 respondents). Out of this sample, respondents who have been retired are 1,402 in number and 1,525 are whose birth years were 1965 and above. This research emphases on if TWs_1 and TWs_2 have an impact on the sub groups estimates which are under analysis (which are retired and had birth years of 1965 and after) and also the estimates on the base of entire sample compared differently with *SWs*. It can be said that we analysis if the S-TWS proposed in this research will assist in adjusting the non-response different than as adjusted in SWA.

To analysis this, we will be conducting two different sets of analysis. First analysis is related to the retired respondents. We will be guesstimating a model to analysis the retired respondents for the elements of psychological well – being. According to Kim & Moen (2002), one of significant social aspects is psychological health which is considered to be also get affected in positive and negative way in later life for example retirements. Hence, it will probably will be more suitable to analyze these weights effect on the retired respondents. As the S-TWS weight set includes weights of sample's two main sub groups, we will be analyzing the impact of S-TWA on the estimations of entire sample. Hence, we have aligned another model for the elements of psychological well-being to be used for entire sample. For both type of analysis (complete sample and retired respondents), we guesstimate the same model by utilizing all three sets of weights which are *SWs*, TWs_1 and TWs_2 individually.

The analysis of other set is related to the ones born in year 1965 and after. For this analysis, we have devised a model to find out the elements of SRM (desire for residential mobility). According to Sadig & Banany (2015), DRM is basically a social phenomenon which states the wishes of individuals to alter their address. Most common between the young individuals is known as residential mobility, it will be significant to evaluate the S-TWA and SWA by analyzing the DRM for the ones born in year 1965 and after (the young ones of the sample). Similarly other analysis will be done related to the retired respondents, we analysis the elements of DRM for the ones born in year 1965 and after and for the complete sample individually. Hence, for both analysis we will guess same model by utilizing all three weight sets *SWs*, TWs_1 and TWs_1 individually.

Similarly, there will be total 12 models to be used in investigation. Out of these, 6 will be used for the psychological well-being analysis and other will be used for DRM analysis. To avoid ambiguity, we have determined the models as mentioned below.

Physiological well-being

Model 1 is projected by using SWs and retired respondents

Model 2 is projected by using TWs_1 and retired respondents

Model 3 is projected by using TWs_2 and retired respondents

Model 4 is projected by using SWs and complete sample

Model 5 is projected by using TWs_1 and complete sample

Model 6 is projected by usingTWs_2 and complete sample

DRM

Model 7 is projected by using SWs and the people born in 1965

Model 8 is projected by using TWs_1 and the people born in 1965

Model 9 is projected by using TWs_2 and the people born in 1965

Model 10 is projected by using SWs and the complete sample

Model 11 is projected by using TWs_1 and the complete sample

Model 12 is projected by using TWs_2 and the complete sample

By using this strategy, a logical comparison can be drawn among these sets of weight which are under analysis for respondents of different sets. This is mainly the same model which have been projected for all sets of weights individually. Hence, differences among corresponding estimates drawn from the use of different weights set which can highlight the differences among the weights as the process of estimation will be constant. This will assist us to define the differences among S-TWA and SWA, however will also allow to make a comparison among the different strategies which have been adopted to establish personalized weights (for both sub groups, interaction based approach and modeling will be used individually).

5.2 Psychological well being

Measures of psychological well-being

Psychological well-being can be measured by using a range of variables available in BHPS. However, the most suitable variables might be the ones which are existing among the GHQ (General Health Questionnaire). According to Taylor, Jenkins & Sacker (2011), this is mainly due to the fact that GHQ variables can be used as consistent measures for psychological well-being. These will be total twelve items and can be acquired by asking the right questions such as mentioned below.

- ✓ Have you been able to work with full concentration recently?
- ✓ Have you lost sleep over something bothering you recently?*
- ✓ Have you been important part in several situations?
- ✓ Have you felt that you are talented enough to make decisions recently?
- ✓ Have you been feeling stress recently?*
- ✓ Have you been able to cope up with challenges recently?*
- ✓ Have you been enjoying your daily activities recently?
- ✓ Have you been able to encounter the challenges recently?
- ✓ Have you been feeling unsatisfied recently?*
- ✓ Have you lost confidence in yourself recently?*
- ✓ Have you been considering yourself insignificant recently?*
- ✓ Have you been happy recently, provided all the things same as usual?

Participants were asked to do rating of all items on the based on four point scale which are more less than usual, less than usual, same as usual and more than usual. To each answers, codes have been given such as 0, 1, 2 & 3 respectively. * Questions have been reversely coded. The GHQ items have been accumulated to establish a basic score which calculates all the cases' mental disturbances in these samples. This is called likert scale or score. The range of this scale is 0 to 36, where low score reflect better feelings of one's well-being and high mental stress rated with high scores. Psychological well-being which is a dependent element which will be calculated through likert scale.

It must be noted that GHQ as mentioned in BHPS is mainly a questionnaire which can be self-completed. Hence, it is probable that the ones who fill the questionnaires in time will reflects good status in terms of health. This might not be the case for all the retired members of the sample (the older participants of the sample) which may have an influence on the estimates drawn from these analysis. However, this challenge is adopted on all sets of weights which are analyzed. Hence, by adopting the modeling approach consistent and making the weights variable, differences among the resultants estimates will be mainly because of the differences among the weighing structures. Hence, this analysis assists in accomplishing the analysis objectives.

In order to test the S-TWS impact on the descriptive statistics, we have divided the likert score into two sub categories determining the better psychological health and worst psychological health[3]. By using *SWs*, TWs_1 and TWs_2 individually, we will measure the proportion of retired respondents in the sample for each category to draw a comparison. The outcome of proportions and linked standards errors are mentioned in table 5.1. This can be noticed from the table that the standard errors of these proportions are closely similar reflecting no significant change related to the effect on the projected precision

[3] Values from 0 to 18 represent good psychological health while values from 19 to 36 indicate bad psychological health.

levels within all sets of weights. However, as TWs_1 and TWs_2 led to same proportions, SWs leads to slightly change proportions (having 1% difference). This outcome reflects that S-TWA will have diverse effect on the extent of the estimate in comparison to SWA.

Table 5.1 Proportions of retired sample members with good and bad psychological health.

	Using *SWs*		Using TWs_1		Using TWs_2	
	%	SE	%	SE	%	SE
Good Ps.health	86	.0029	87	.0028	87	.0029
Bad Ps.health	14	.0029	13	.0028	13	.0029

* Ps.health is psychological health. SE is the standard error.

Modeling psychological well-being

In likert score analysis, the outcome variable is a continuous variable. The data structure (which is based on the numerous observations of each person) permits the use of panel models of data. Hence, we have projected OLS regression model's random impacts to analysis the psychological well-being's elements of the retired participants and the complete sample individually. As mentioned earlier, each model was projected three times by altering the weights among *SWs*, TWs_1 andTWs_2. According to Taylor, Jenkins & Sacker, 2011; Ryan & Frederic, 2006; Kohler, Behman & Skytthe, 2005), psychological well-being is related to the calculations of health, ethnicity, wealth, age and cohabitation. Hence, the variables applied for psychological well-being which is selected in parallel to these processes. These variables are for instance, age and race where respondents is living with his/her partner, health status, income and savings. Furthermore, rest of the variables, for instance gender and time will be used in control model.

Before having a discussion about modeling psychological well-being's outcome, we have to devise a criteria for determining the differences among the estimates of both substantive analysis (DRM and psychological well-being). We have opted the same method of testing hypotheses (incorporating confidence intervals) which will be adopted for substantive analysis as mentioned in chapter 2 to determine the important difference among the equivalent estimates adjusted with different weights. To avoid ambiguity, this approach is explained again below.

To determine the important differences among the equivalent coefficients projected with different sets of weights, we proceeded for hypotheses testing on the differences among the equivalent estimates accustomed with S-TWA and SWA by using 95% of CI (confidence intervals). Test will be done in two main steps. First step is all about establishing 95% of CIs of the difference among both equivalent coefficients which are accustomed with S-TWA and SWA. These CIs will determine the extent of values about the differences among two equivalents coefficients may resides. For instance, ifβ_{SW}, β_{TW1} and β_{TW2} signify the provided set of equivalent population factors projected by equivalent coefficient setb_{SW}, b_{TW1} and b_{TW2} managed with *SWs*, TWs_1 and TWs_2 separately, we establish two CIs to analyze if b_{SW} is unlike than b_{TW1} and b_{TW2}. CIs for these will be (b_{SW} - b_{TW1}) and (b_{SW} - b_{TW2}). These CIs are 95% and are provided in below 5.1.

$(b_{SW} - b_{TWi}) \pm 1.96 * S_{b_{SW}-b_{TW}}$ (5.1).

Where standard error is $S_{b_{SW}-b_{TWi}}$ of $(b_{SW} - b_{TWi})$ and is provided in 5.2 and i=1, 2.

$$S_{b_{SW}-b_{TWi}} = \sqrt{S^2(b_{SW}) + S^2(b_{TWi}) - 2 * Cov(b_{SW}, b_{TWi})} \quad (5.2)$$

Where variance of b_{SW} and b_{TWi} are $S^2(b_{SW})$ and $S^2(b_{TWi})$ respectively; and covariance of b_{SW} and b_{TWi} are $Cov(b_{SW}, b_{TWi})$; and i=1, 2.

The next step is to adopt the established CIs to test if there is a major difference among both equivalent coefficients accustomed with S-TWA and SWA (for instance is there a major difference amongb_{SW} and b_{TWi}? This is about testing about below hypothesis.

H_0: $\beta_{SW} - \beta_{TWi} = 0$ againstH_a: $\beta_{SW} - \beta_{TWi} \neq 0$; *i=1,2.*

Where H_0 is rejected where there is a major difference amongb_{SW} and b_{TWi}, i=1,2) when the necessary CI will not be 0.

Same test will be adopted for both substantive analysis (DRM and psychological well-being) and we show 95% CIs for difference among both equivalent estimates (CI of [b_{SW} - b_{TW1}] and CI of [b_{SW} - b_{TW2}]) in necessary outcome table.

Table 5.2 and 5.3 show the outcome of modeling psychological well-being for the retired participants and the entire sample respectively. Initiating with the outcomes from the model of retired respondents (as mentioned in table 5.2), it can be observed that all equivalent estimates within all the models have same level of importance. The outcome is constant with the previous observation that have same contribution on the base of dispersion.

Furthermore, we have instantly observe, as projected that TWs_1 and TWs_2 will lead to same estimates. Majority of the coefficients leading to the outcome from two sets of weights are almost equal. And related to the coefficients outcome from *SWs*, these estimates are same to the estimates resulted from TWs_1 and TWs_2. Nonetheless, coefficient having better condition of health which resulted from *SWs* proves to be importantly unlike from the equivalent coefficients projected with TWs_1 and TWs_2. The variance is determined by two CIs is of the respective difference among (b_{SW} - b_{TW1}) and (b_{SW} - b_{TW2}), where b represents coefficient of better condition health. Both of these CIs will not be 0 determining that the estimate which is to be questioned (accustomed with SWs) is majorly different as compared to the equivalent estimates accustomed with TWs_1 and TWs_2. Concentrating on the models which will be used for complete sample (as mentioned in table 5.3), the outcome does not present proof of important differences among the estimates outcomes from the SWs and their equivalent estimates established with TWs_1 and TWs_2. As recommended from the outcome, the S-TWA and SWA might lead to major difference in outcome in regard to the estimates established from sub-groups which have been selected from custom made weighting. Zero will be included for all the CIs of the variance among (b_{SW} - b_{TW1}) and (b_{SW} - b_{TW2}). However we can observe that coefficient of better health condition which is estimated in the model with

SWs is quite different in comparison to models of equivalent coefficients estimated along with TWs_1 and TWs_2. The CI test conducted do not recommend that the variance is important, such variances only matter while interpreting the outcome of analysis and are significant for altering and updating the understanding related to social phenomena. Hence differences must not be considered insignificant.

Table 5.2 Random effects OLS regression models of the determinants of psychological well-being for retired respondents.

	Using *SWs*	**Using TWs_1**	**95% CI of (b_{SW} - b_{TW1})**		**UsingTWs_2**	**95% CI of (b_{SW} - b_{TW2})**	
Years 1995 to 1998	0.131	0.030	-0.147	0.349	0.084	-0.202	0.296
Female	0.841**	0.884**	-0.351	0.256	0.861**	-0.626	0.586
White	-0.979	-1.268	-2.245	2.832	-1.766	-2.057	3.631
Age	0.020	0.022	-0.020	0.016	0.022	-0.038	0.034
Living with a partner	-1.027***	-1.042***	-0.452	0.482	-1.103***	-0.392	0.544
Has savings	-0.135**	-0.335**	-0.073	0.473	-0.322**	-0.086	0.460
Has a good health condition	-0.726*** [a]	-1.607*** [a]	0.321	1.441	-1.605*** [a]	0.313	1.445
Income/1000	-0.003	-0.004	-0.042	0.026	-0.004	-0.024	0.026
n	1,402	1,402			1,402		
σ	2.79	2.89			2.89		
P	0.46	0.47			0.47		

*All models are estimated by using a balanced panel of retired sample members who responded in the first 8 waves. The reference categories of the independent variables are: years 1991 to 1994, male, non-white, does not live with a partner, has no savings and has a bad health condition. [a] indicates a significant difference between the equivalent estimates adjusted with the *SWs* and both sets of *TWs*. * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

Table 5.3 Random effects OLS regression models of the determinants of psychological well-being for the whole

	Using *SWs*	**Using TWs_1**	**95% CI of (b_{SW} - b_{TW1})**		**UsingTWs_2**	**95% CI**
Years 1995 to 1998	0.262	0.213	-0.067	0.165	0.248	- 0.0
Female	1.217***	1.233***	-0.252	0.220	1.226***	-0.2
White	-0.580**	-0.575**	-0.519	0.509	-0.631**	-0.46
Age	0.003*	0.006*	-0.009	0.003	0.005*	-0.0
Living with a partner	-0.482***	-0.485***	-0.180	0.186	-0.504***	-0.16
Has savings	-0.225***	-0.223***	-0.127	0.123	-0.228***	-0.12
Has a good health condition	-0.955***	-1.141***	-0.453	0.825	-1.162***	-0.39
Income/1000	-0.001	-0.002	-0.006	0.008	-0.002	-0.00
n	6,753	6,753			6,753	
σ	2.92	2.93			2.93	
P	0.36	0.37			0.37	

*All models are estimated by using a balanced panel of those who responded in the first 8 waves. The reference categories of the indep
1991 to 1994, male, non-white, does not live with a partner, has no savings and has a bad health condition. * $p< 0.05$, ** $p< 0.01$, *** $p<$

5.3 Desire for residential mobility (DRM)

Measure of DRM

For social surveys, it is not common practice to add a direct question regarding if participants which to alter their addresses. In BHPS, the participants are always asked a same question that if they would like to move from one place to another. For resultant variable, we will be utilizing this item. Furthermore, when the participants shared their interest of moving to new house, this is considered as a desire from his side for residential mobility. Hence, for this part of our analysis, the dependent variable will be the binary variable, showing if DRM is existing in participants or not. Equation 5.3 describes this variable.

$$DRM_i = \begin{cases} 1, & \textit{if respondent } i \textit{ has intention to move house} \\ 0, & \textit{otherwise} \end{cases} \tag{5.3}$$

DRM modeling

For modeling of DRM, we will be using logistic regression's random effects. This was applied for the ones having year 1965 as their birth year and after and for complete sample individually. For all participants' set, estimated model was used along with all sets of weights which are under analysis. For this analysis, the independent variables will be gender, household size, and age, and race, number of rooms of house, housing tenure and savings possessions. The selection of these variables are on the base of residential mobility's literature (Sanbonmatsu et al, 2011) and the accessibility of these variable related to analysis all 8 waves.

DRM modeling outcomes have been shown table 5.4 for those participants born in year 1965 and after and for complete sample, outcomes are mentioned in table 5.5. You can observe that both tables are showing outcomes of odd ratios. As earlier proceeded for psychological well-being's analysis, 95% CIs will be used to differentiate the variance among the estimates established with *SWs* (b_{SW}) and their equivalent estimates established with TWs_1 (b_{TW1}) and TWs_2 (b_{TW2}) to test whether two S-TWS and SWA will outcome to different estimates.

Concentrating on the model for participants born in year 1965 and after (please refer to table 5.4), we have observed that the outcome estimates are same within all these models. Keeping the similar significance levels, majority of the coefficients are same on the base of magnitude. More estimates are outcomes from TWs_1 and TWs_2 as compared to estimates outcomes from SWs and both of the costumed weights.

However, SWs generated one main difference in comparison to TWs_1 andTWs_2. Mainly, 'member of large household' coefficient seemed to be more importantly different as compared to TWs_1 and TWs_2 estimated coefficients. This has been backed by two CIs of the variance among the estimate and its equivalent estimates which are established by TWs_1 and TWs_2. 0 will not be included in both CIs. This outcome is aligned with the outcomes from the participants who are retired according to the model recommending that S-TWS might lead to outcome in some differences as compared to the SWA, mainly when we limit the analysis to sub-groups utilized to establish the tailored weights.

Moving to the models related to the complete samples (please refer to table 5.5), the outcome has not important variances within the estimates. On the base of 95% CIs test, it has been noticed that all CIs of the differences among the estimates managed in SWs and their equivalent estimates managed with TWs_1 and TWs_2 accordingly showing zero as a value of difference showing that both equivalent estimates are not dominatingly different. Nonetheless, alike psychological well-being analysis, we noticed a variance among the 'member of large household' coefficient in this model projected with SWs and TWs_1 and TWs_2 estimated equivalent estimates. As mentioned earlier, these differences are of significant importance and they show that S-TWA might have an impact on the estimates of complete sample. On the base of outcome, it might be rational to assume important variances among the complete sample estimating the outcome from S-TWA and SWA when other sub-group are used for the tailored weights of sub-group.

In short, on the base of this analysis, the outcome recommends that weightings both approaches (SWA and S-TWA) are alike in terms of the complete impact on the estimates. Nonetheless, S-TWA might lead to different effect on few estimates. On the base of CIs test, these variances will be proved to be important. Considering the S-TWA estimated weighting models, indicates the sub-groups response process under analysis effective as compared to the SWA weighting models, estimates which turns to be difference using S-TWA which are not biased to same extent as SWA generated equivalent estimates.

Moreover, both sub-group's approaches related to tailored weighting (modeling response and interaction based used individually for all sub-groups) appear to be analogous in reference to the outcome weights. Weights do not appear to have a different effect on the estimates. Moreover, the S-TWA have an effect on the sub-groups' established estimates which is in question, in comparison to the estimates established from analysis of complete sample. In the later scenario, though S-TWA has an outcome of few rational differences, out analysis depicts that these modifications might not be important as the ones used from analysis limited to the S-TWA selected sub-groups. It is depicting from these differences that S-TWS might have an influence on to complete sample estimate, and for other sub-groups, it could be a possibility to prove that this impact can lead to the outcome of important difference.

Table 5.5 Random effects logistic regression models of the determinants of the desire for residential mobility for those who were born in 1965 or after.

	Using *SWs*	**Using TWs_1**	**95% CI of (b_{SW} - b_{TW1})**		**Using TWs_2**	**95% CI of (b_{SW} - b_{TW2})**	
Years 1995 to 1998	1.137	1.175	-0.168	0.092	1.170	-0.163	0.097
Female	0.866*	0.868*	-0.220	0.126	0.864*	-0.216	0.220
White	1.041	1.081	-0.793	0.713	1.021	-0.320	0.360
Age	1.045	1.044	-0.003	0.005	1.044	-0.003	0.005
Member of a large household	1.057* [a]	1.133* [a]	-0.126	-0.026	1.136* [a]	-0.130	-0.028
Lives in a house with 3 to 4 rooms	1.242*	1.255*	-0.246	0.220	1.251*	-0.242	0.224
Lives in a house with 5+ rooms	1.365***	1.386***	-0.285	0.244	1.377***	-0.276	0.252
Has savings	0.979	0.971	-0.088	0.104	0.970	-0.087	0.087
House owned outright	0.845*	0.826*	-0.180	0.218	0.838*	-0.192	0.192
House owned with mortgage	1.562***	1.560***	-0.221	0.225	1.532***	-0.583	0.643
n	1,525	1,525			1,525		
σ	1.44	1.45			1.45		
P	0.38	0.39			0.39		

* The entries are odds ratios. All models are estimated by using a balanced panel of those who were born in 1965 or after and who responded in the first 8 waves. The reference categories of the independent variables are: years 1991 to 1994, male, non-white, member of a small household (3 members or less), lives in a house with 1 or 2 rooms, has no savings and tenant. [a] indicates a significant difference between the equivalent estimates adjusted with the *SWs* and both sets of *TWs*.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.5 Random effects logistic regression models of the determinants of the desire for residential mobility for

	Using *SWs*	Using TWs_1	95% CI of (b_{SW} - b_{TW1})		Using TWs_2
Years 1995 to 1998	1.165	1.164	-0.048	0.049	1.164
Female	1.180*	1.171*	-0.120	0.138	1.181*
White	1.241*	1.243*	-0.313	0.309	1.202*
Age	0.964*	0.972*	-0.046	0.030	0.973*
Member of a large household	1.053**	1.107**	-0.130	0.022	1.109**
Lives in a house with 3 to 4 rooms	1.067*	1.068*	-0.132	0.130	1.083*
Lives in a house with 5+ rooms	1.191**	1.195**	-0.153	0.145	1.207**
Has savings	1.054	1.049	-0.053	0.063	1.53
House owned outright	0.836***	0.854***	-0.314	0.278	0.869***
House owned with mortgage	1.587***	1.569***	-0.125	0.161	1.579***
n	6,753	6,753			6,753
σ	2.25	2.27			2.26
P	0.60	0.61			0.61

* The entries are odds ratios. All models are estimated by using a balanced panel of those who responded in the first 8 waves. The independent variables are: years 1991 to 1994, male, non-white, member of a small household (3 members or less), lives in a house savings and tenant. * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

5.4 Discussion

In this chapter, we have analysis alternate approach of weighting (the tailored weighting of sub-group). The S-TWA (the sub-group tailored weighting) is on the base of choosing particular sub-groups from survey sample and tailoring the sub-groups non-response weight construction. Contrasting to SWA, on the base of S-TWA, the weights are established by utilizing the weighting variables set that have an impact on the response probability in the chosen sub-group irrespective of, if or not it will have an impact on the response probability in the balance sample. Also, the S-TWA estimated weighting models may have limit to the members of the sample from all the sub-groups. Furthermore, both approaches to conduct S-TWA is applied, modeling the response propensity and interaction-based approach individually for sub-group.

Chapter's main findings have been summarized as below

1. S-TWA impact on the estimates which are usually same as SWA, in certain cases related to estimates precision levels.
2. On few estimates, the S-TWA generated different outcomes (on the base of magnitude) as compared to SWA.
3. It appears likeable for S-TWA to impact all sample estimates and sub-groups derived estimates chose for tailored weighting. Nonetheless, the impact appears to be important for the estimates established from the tailored weighting of the sub-groups.

4. Both approaches introduced of S-TWS seems to generate same set of weighting that is resulted from the estimates similar effects.

These outcomes concentrates on the numerous propositions in longitudinal surveys, for the establishment of non-response weighting. Firstly, the outcome recommends that different weights sets which are generated from the S-TWS is more or less difference as compared to the weights set that resulted from SWA. The variance is noticed as an outcome of different methodology which have utilized to establish the tailored weights. Altering the standard covariates (non-response) and limiting the sub-groups weighting model for which tailored weights are established which lead to a results in tailored weights set which have different values of weights as compared to standard weights. As an outcome, the tailored weights might lead from few estimates to distinguish from the equivalent estimates established from standard weights. If non-response covariates both changes and respondents set used in S-TWA depicts the sub-groups' non-response process which is effective than SWA, S-TWA is considered to manage the sub-groups non-response under analysis significant than SWA.

Second, although our investigation here does not show evidence that the S-TWA results in significantly different estimates than the SWA when estimates are derived from full sample analysis, it shows that some of the total sample estimates may still change considerably in terms of their magnitude if adjusted with the S-TWA. We believe that such changes, sometimes, have different impact on the interpretation of the results, especially with sensitive measures in some of the socio-economic processes. Hence, different conclusions regarding some of the total sample estimates could still be drawn on the basis of the S-TWA.

Third, our analysis suggests that the two approaches of the subgroup-tailored weighting (*interaction-based approach and modeling the response propensity separately for each selected sub-group*) may substitute one another. However, one may still expect differences – maybe not to a large extent - between these two approaches if they are applied on a different data set or different sub-groups. This is especially so if the number of the proposed variables for the tailored weighting is large. Thus, if the S-TWA is considered, we recommend the application of the second approach (*modeling the response propensity separately for each selected sub-group*) because it has some advantages over the first one (*interaction-based approach*). One of these advantages is that the second approach avoids the complications associated with too many interactions in the weighting model. Another advantage is that it allows restricting the weighting model to sample members in the sub-group selected for tailored weighting, which in turn permits excluding variables that do not predict response in the sub-group in question.

The availability of a large number of weighting variables in longitudinal surveys is advantageous. However, any weighting approach that depends on using a large number of variables to model the response propensity in the sample, but assumes that the effects of these variables are the same for different sub-groups (such as the SWA), may not always explain the non-response process well in all sub-groups in the sample. This is because samples in longitudinal surveys are large, and are often composed of units from a number of sub-populations which are not necessarily homogeneous in terms of the factors responsible for non-response. Successful weighting, in our opinion, depends on an independent and profound understanding of the non-response process in each of the major sub-groups in the sample rather than the number of variables included in a single weighting model. Even in the same survey sample, the cause of non-response may differ vastly across some sub-groups suggesting different sets of weighting variables (both in terms of scale and type) for weighting. Thus, looking at the non-response reasons in the sample as a whole may lead to ignoring variables that may appear insignificant in general while they are in fact important to explain non-response in some sub-groups. The findings in this chapter have demonstrated this. For example, it is known that factors like 'age' are powerful weighting variable while factors such as 'religion' are weak predictors of non-response; though, the results of this investigation showed the exact opposite within the subgroups on which we have focused. At first glance, it may be hard to understand how a – well known - powerful auxiliary as 'age' could not be important in predicting response while a variable such as 'religion' is significant. However, once the cause of non-response is understood at a sub-group level, it can all be explained.

We expect similar findings if the S-TWA is applied in other panel studies, such as Understanding Society for example. However, in such a large longitudinal survey, the application of S-TWA might be, to some extent, tricky. This is because identifying the number of sub-groups that the tailored weighting should be based on is a subjective matter. S-TWA may be more appropriate for specific analyses where the analyst wants the best possible weights for a specific purpose. As for general-purpose public-release weights, it could be challenging to produce the best possible set of tailored weights because, in longitudinal survey samples, sub-groups maybe identified in a number of dimensions. Therefore, it would be difficult to identify a number of sub-groups that allows the execution of the best subgroup-tailored weighting. However, it should be pointed out that the more sub-groups used to create tailored weights (bearing in mind that the relevant sample sizes should be large enough to estimate non-response well) the stronger the effect of the overall set of tailored weights will be.

Additionally, even if the number of the required sub-groups is accurately identified, the survey organization will face the problem of identifying "which specific sub-groups should be used for tailored weighting?" as this maybe a subjective matter too.

Sub-groups can be non-overlapping (e.g. the sub-groups used in the analysis of this chapter). In this case, the sub-sets of tailored weights can be put together to form an Overall Set of Tailored Weights (OSTW). Accordingly, the OSTW can be beneficial in analyses that target the whole sample or analyses restricted to sub-groups. However, the sub-groups selected for tailored weighting maybe overlapping (e.g. sub-groups of males, disabled and white respondents). In this case, producing a OSTW is not possible via this method and, therefore, a number of sets of tailored weights may need to be released separately. However, this type of sub-group tailored weighting may not be appropriate for total sample estimates adjustment.

Thus, a future research may investigate a procedure that decides on:

- The number of sub-groups required for an effective overall set of tailored weights.
- Whether sub-groups should or should not be overlapped.
- Which specific sub-group should be selected for tailored weighting.

Finally, as the S-TWA uses a different set of variables compared to the one used in the SWA, researchers who are deciding between tailored weights and standard weights should pay attention to the set of variables used to create the tailored weights. This is because weights are also powerful in dealing with non-response bias if they are created using a set of variables that is strongly correlated with the main variable in the analysis (the dependent variable). Therefore, standard weights may also be a good choice if its weighting variables are more correlated with the dependent variable in the analysis. In this case it is a tradeoff between the reward of the tailored weights and the relationship between the dependent variable and the weighting variables used to create the standard weights. Thus, if a survey organization considered S-TWA as an alternative, it may still want to keep standard weights in the public data files. Moreover, if a set of tailored weights is included in the data files, the survey organization should properly document the process of weights creation as well as clearly stating the variables used to create the weights.

Chapter 6

Conclusion, Recommendations

6.1 Conclusions:

The major findings of this thesis can be summarized in four main points:

1- The effect of the S-TWA on estimates is generally similar to that of the SWA, in particular in terms of estimates precision levels.

2- On some estimates, the S-TWA produces different results (in terms of magnitude) than the SWA.

3- It seems possible for the S-TWA to affect both total sample estimates and estimates derived only from the sub-groups selected for tailored weighting. However, the effect seems to be stronger (significant) on the estimates constructed from the sub-groups selected for the tailored weighting.

4- The two introduced approaches of S-TWA appear to produce similar sets of tailored weights that result in the same effect on estimates.

This research focused on unit non-response weights in longitudinal surveys.

Usually theses weights are built using the standard weighting approach (SWA).The (SWA) suffers some limitations and this thesis attempted to overcome these.

Most major longitudinal surveys in the world today construct unit non-response weights using (SWA) which ignores variables that are considered important only at sub-groups levels. In this thesis it was shown that adopting (SWA) may lead to biased estimates with low precision.

As a result, a new approach (S-TWA) was suggested by the researcher which was based on taking important variables for members of the sample in certain sub-groups.

A main finding of this research is that weights arrived at by (SWA) may not be appropriate when used for estimation at a sub-group level.

It was also shown that when some sub-groups in the sample have important variables that are not the same as variables for the whole sample the S-TWA leads to less biased and more precise as compared to the SWA. This may be due to the fact that expectation of response to the questionnaire on these variables in the sub-groups is higher than the response for the whole sample. However when no subgroups have specific important variables the approaches lead, in general, to similar results.

Another result revealed that the standard errors of estimates using (S-TWA) are, in all cases, lower than when using (SWA).

6.2 Recommendations:

In light of the above results the following recommendations may be made:

1- Data analysts who use weighting in their analyses are recommended to recognize the problem referred to in this research and use (S-TWA) beside the (SWA) especially when it is known that some sub-groups in the sample have specific variables important to weighting.

2- Users of the weights provided by the organizations should not assume that they are automatically applicable to all members in the sample. If it turned out that some sub-groups in the sample have important variables different form those used in deriving the weights by (SWA) the analyst should use S-TWA for deriving appropriate weights.

3- Survey organizations in Sudan should review their weighting systems and attempt to implement the tailored weighting approach for more development in the non-response weighting.

References

Allison, P. (2000) “Missing Data”. Thousand Oaks, CA: Sage.

Atrostic, B.K, Bates, N., Burt, G., and Silberstein, A. (2001) “Non-response in U.S. Government Household Surveys: Consistent Measures, Recent Trends and New Insights”, *Journal of Official Statistics*, Vol. 17, pp. 209-226.

Becketti, S., Gould, W., Lillard, L., and Welch, F. (1988) “The Panel Study of Income Dynamics after fourteen years: an evaluation”, *Journal of Labour Economics,* Vol. 6, pp. 472-492.

Bethlehem, J., and H. Kersten (1986): \Werken met Nonrespons (in Dutch)," Ph.D. thesis, *Statistische Onderzoekingen* M30.

Bethlehem, J. (1988): \Reduction of Nonresponse Bias through Regression Estimation," *Journal of O±cial Statistics, 4(3), 251{260*.

Bethlehem, (1999): \Cross-sectional Research.," in *Research Methodology in the Social, Behavioural & Life Science,* ed. by H. Adµer, and G. Mellenbergh, pp. 110{142. Sage Publications, London.

Bethlehem, J., and B. Schouten (2004): \Nonresponse Analysis of the Inte-grated Survey on Living Conditions (POLS)," Discussion paper 04004, Statis-tics Netherlands, Available at www.cbs.nl.

Bethlehem, J.(2009) “Applied Survey Methods: A Statistical Perspective”, New Jersey: John Wiley& Sons, Inc.

Bethlehem, Jelke; Cobben, Fannie; and Schouten, Barry, Handbook of Nonresponse in Household Surveys. New York: John Wiley & Sons, 2011.

Biemer, P., and L. Lyberg (2003*): Introduction to Survey Quality*. John Wiley & Sons, Inc.

Biemer, P. P., and Christ, S. L. (2008) "Weighting Survey Data", (chapter 17) in De Leeuw, E.D., Hox, J.J., and Dillman, D.A., *International Handbook of Survey Methodology*. New York: Lawrence Erlbaum.

Brandon, E. (2012) "Why Older Citizens are More Likely to Vote".[online] available from http://money.usnews.com/money/retirement/articles/2012/03/19/why-older-citizens-are-more-likely-to-vote [Accessed 23d June 2013].

Bradburn, Norman M. 1992. ' Presidential Address: A Response to the Nonresponse Problem. ' Public Opinion Quarterly 56: 391–397.

Brehm, J. (1993)"The phantom respondents: Opinion surveys and political representation". Ann Arbor: University of Michigan Press.

Brick, J. (2013) "Nonresponse and Weighting Adjustments: A critical review", *Journal of Official Statistics*, Vol. 29(3), pp. 329-353.

Carpenter, J., and Plewis, I. (2011) "Analysing Longitudinal Studies with Non-response: Issues and Statistical Methods", (pp. 518-543) in Williams, M., and Vogt, W., *Hand Book of Methodological Innovation.* Newbury Park: Sage.

Chang, M. (2010) "Adjusting for Non-response Bias Due to Attrition and Intermittent Wave Non-response in Longitudinal Surveys of Older Adults",*Dissertation*. Nebraska: University of Nebraska.

Cochran, W. (1977) “Sampling Techniques”, John Wiley and Sons.

Cobben, Fannie, Nonresponse in Sample Surveys: Methods for Analysis and Adjustment. The Hague: Statistics Netherlands, 2009.

Couper, M. P. (1997)“Survey introductions and data quality”, *Public Opinion Quarterly*,Vol. 6, pp. 317-338.

David, M., Little, R. J. A., Samuhel, M. E., and Triest, R. K. (1983) “Non-random Non-response Models Based on the Propensity to Respond”, Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp. 168-173.

de Leeuw, Edith and Wim de Heer. 2003. ‘Trends in Household Survey Non-response:A Longitudinal and International Comparison.’ Pp. 41–54 *in Survey Non-response,* edited by Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little.New York: John Wiley & Sons.

Dillman, Don A., John L. Eltinge, Robert M. Groves and Roderick J.A. Little. 2002.

‘Survey Non-response in Design, Data Collection, and Analysis.’ Pp. 3–26 *in Survey Non-response*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J.A. Little. New York: John Wiley & Sons.

Duncan, G., Hofferth, S. L., and Stafford, F. P. (2004) “Evolution and Change in Family Income, Wealth and Health: The Panel Study of Income Dynamics, 1968-2000 and beyond”. (chapter 6) In House, J. (Eds), *A telescope on Society*. Ann Arbor: Institute for Social Research, University of Michigan.

Fitzgerald, J., Gottschalk, P., and Moffitt, R. (1998)"An Analysis of Sample Attrition in Panel Data: the Michigan Panel Study of Income Dynamics", Journal of Human Resources, pp.251-299.

Goldstein, H. (2009) "Handling Attrition and Non-response in Longitudinal Data", *Longitudinal and Life Course Studies*, Vol. 1, pp. 63-72.

Goyder, J. (1985) "Face-to-Face Interviews and Mail Questionnaires: The Net Difference in Response Rate", *Public Opinion Quarterly,* Vol. 49, pp. 234-252.

Groves, R., and Couper, M. (1998) "Nonresponse in Household Interview Surveys", New York: Wiley.

Groves, R., Presser, S., and Dipko, S. (2004) "The Role of Topic Interest In Survey Participation Decisions", *Public Opinion Quarterly*, Vol. 68, pp. 2-31.

Groves, R., Fowler,F., Couper, M., Lepkowski, J., Singer, E., and Tourangeau, R.(2004)*"Survey Methodology",* New York: Wiley.

Groves, R., Wagner, J., and Peytcheva, E. (2007) "Use of Interviewer Judgments About Attributes of Selected Respondents in Post-survey Adjustment for Unit Non-response: An Illustration With the National Survey of Family Growth", Proceedings of the Survey Research Methods Section, ASA.

Hansen, M., and W. Hurwitz (1946): \The Problem of Non-response in Sam-
ple Surveys," *Journal of the American Statistical Association,* 41, 517{529.

Hausman, J. A., and Wise, D. A. (1979)"Attrition bias in experimental and panel data: The Gary income maintenance experiment", *Econometrika* , Vol.47, pp. 455-473.

Heckman, J. (1979) “Sample Selection Bias as a Specification Error”. *Econometrica*, Vol. 47, pp. 153-162.

Horvitz, D. G., and Tompson, D. J. (1952) “A generalisation of Sampling without Replacement from a Finite Universe” *Journal of the American Statistical Association*, Vol (47), pp. 663-685.

Kalton, G., and I. Flores-Cervantes. (2003). “Weighting Methods.”, *Journal of Official Statistics*, Vol. 19, pp. 81-97.

Kish, L. (1965) “Survey Sampling”, New York: John Wiley and Sons.

Kim, J., and Moen, P. (2002) “Retirement Transitions, Gender, and Psychological Well-being: A life-course, Ecological Model”, Journal of Gerontology, Vol. 57B(3), pp. 212-222.

Kohler, H., Behrman, J., and Skytthe, A. (2005) “Partner + Children = Happiness? The Effects of Partnerships and Fertility on Well-Being” *Population and development review*, Vol. 31, issue 3, pp.407-445.

Kreuter, F., Lemay, M., and Casas-Cordero, C. (2007) “Using Proxy Measures of Survey Outcomes in Post-survey Adjustments: Examples From the European Social Survey (ESS)”, Proceedings of the Survey Research Methods Section, ASA.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R., and Raghunathan, T.(2010) “Using Proxy Measuresand Other Correlates of Survey Outcomes to Adjust for Nonresponse: Examples From Multiple Surveys”, Journal of the Royal Statistical Society, SeriesA, Vol. 173, pp.389-407.

Kreuter, F., and Olson, K. (2011) “Multiple Auxiliary Variables in Non-response Adjustment”, Sociological Methods & Research, Vol.40, pp. 311–332.

Kroh, M. (2009) “Data Documentation: sample sizes and panel attrition in the German Socio Economic Panel (1984 until 2008)”, Berlin: Deutsches Institut für Wirt schafts for schung (DIW).

Laurie, H., and Lynn, P. (2009) “The Use of Respondent Incentives on Longitudinal Surveys”, in Lynn, P., *Methodology of Longitudinal Surveys.* Sussex: John Wiley & Sons, Ltd.

Lepkowski, J., and Couper, M. (2002) “Non-response in the Second Wave of Longitudinal Household Surveys”, in Groves, R., Dillman, D., Eltinge, J., and Little, R., *Survey Nonresponse*. New York: John Wiley.

Lessler, J., and Kalsbeek, W. (1992) “Nonsampling Error in Surveys”, New York, John Wiley & Sons, Inc.

Little, R. (1986) “Survey Nonresponse Adjustments for Estimates of Means.” *International Statistical Review,* Vol. 54, pp.139-57.

Little, R.J., & Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* New York: John Wiley & Sons.

Little, R. J. A., and Rubin, D. B. (2002) “Statistical Analysis with Missing Data”. New York: Wiley.

Little, R. J., and Vartivarian, S. (2003) “On Weighting the Rates in Non-response Weights”, Statistics in Medicine, Vol. 22, pp.1589-599.

Little, R. J., and Vartivarian, S. (2005) “Does Weighting for Nonresponse Increase the Variance of Survey Means?”,Survey Methodology, Vol.31, pp.161-68.

Lynn, P. (1996) “Weighting for Non-response”, pp. 205-214 in Banks. R. et al. Survey and statistical computing 1996, ASC.

Lynn, P. (2005) “Weighting”, pp.967-973 in Kempf-Leonard. K. *Encyclopedia of Social Measurment*, Academic press.

Lynn, P. (ed.). (2006) “Quality Profile: British Household Panel Survey. Version 2.0: Waves 1to 13: 1991-2003”. Colchester: Institute for Social and Economic Research, University of Essex. [online] available from http://www.iser.essex.ac.uk/ulsc/bhps/quality-profiles/BHPS-QP-01-03-06-v2.pdf[Accessed 4th November 2010].

Lynn, P. (2008) “The Problem of Non-response”,(pp. 35-55) in De Leeuw, E.D., Hox, J.J., and Dillman, D.A., International Handbook of Survey Methodology. New York: Lawrence Erlbaum.

Lynn, P. (2009) “Methods for Longitudinal Surveys”, (chapter 1) in Lynn, P., *Methodology of Longitudinal Surveys.* Sussex: John Wiley & Sons, Ltd.

Lynn, P. (2011) “Maintaining Cross-Sectional Representativeness in a Longitudinal General Population Survey”, *Institute for Social and Economic Research Working Papers*, No. 2011-04. Colchester: University of Essex.

Lynn, P., Kaminska, O., and Goldstein, H. (2014) “Panel Attrition: How Important is it to Keep the Same Interviewer?”, *Journal of official statistics*, Vol. 30, pp.443-457.

Plaza, S., and Graf, E. (2008) “Executive Summary: Recommendations and Practical Examples for Using Weighting”, Lausanne: University of Lausanne.

Plewis, I. (2011) “Auxiliary Variables and Adjustments for Missingness in Longitudinal Studies”, *the Cathie March Centre for Census and Survey Research*, eScholar ID: 138810. Manchester: University of Manchester.

Rubin D.B. (1976), “Inference and Missing Data”, Biometrika, 63: 581—592.

Ryan,R., and Frederick,C. (2006) “On Energy, Personality, and Health: subjective vitality as a dynamic reflection of well-being”, *Journal of personality*, Vol. 65, issue 3, pp.529-565.

Rosenbaum, P. R., and Rubin, D. B. (1983) “The Central Role of the Propensity Score in Observational Studies for Casual Effects”, Biometrika, Vol. 70, pp. 41-55.

Sadig, H. (2011). Non-response Weight Adjustments in Longitudinal Surveys. In *Survey Research Methods and Applications*, (pp.207-210). Pisa: Edizioni plus-Pisa University Press.

Sadig, H. (2014) “Non-response Subgroup Tailored Weighting: the Choice of Variables and the Set of Respondents Used to Estimate the Weighting Model”, *Institute for Social and Economic Research Working Papers*, No. 2014-36. Colchester: University of Essex.

Sadig, H., and Banany, A. (2015) “The Impact of Non-response Weighting: Empirical Evidence From Modelling Residential Mobility”, *Journal of Social Research and Policy*, (Forthcoming).

Sanbonmatsu, L., Ludwig, J., Katz, L. F., Gennetian, L. A., Duncan, G. J., Kessler, R. C. (2011) “Moving to Opportunity for Fair Housing Demonstration Program - Final Impacts Evaluation”. Cambridge, MA: National Bureau of Economic Research.

SÄarndal, C., B. Swensson, and J. Wretman (1992): *Model Assisted Survey Sampling.* Springer-Verlag, New York.

Särndal, C. E., and Lundström, S. (2005) “Estimation in Surveys with Nonresponse”, Chichester: John Wiley & Sons Ltd.

Schafer JL (1997). Analysis of incomplete multivariate Data. London: Chapman and Hal .

Schafer, J., and J. Graham (2002): \Missing Data: Our View of the State

of the Art," *Psychological Methods*, 7(2), 147{177.

Schouten, B. (2004) “Adjustment for Bias in the Integrated Survey on Household Living Conditions (POLS) 1998”, *Statistics Netherlands Discussion Paper,* NO. 04001. Voorburg: Statistics Netherlands.

Seal,K.C.(1962).Use of outdated farms in large scale sample surveys. *Calculatta Statist.*Assoc.Bull.11.

Siddiqui, O., Flay, B., and Hu, F. (1996)“Factors Affecting Attrition in a Longitudinal Smoking Prevention Study”, *Preventive Medicine,*Vol. 25**,**pp. 554-560.

Sikkel, D., Hox, J., and De Leeuw, E. (2009) “Using Auxiliary Data for Adjustment in Longitudinal Research”. (chapter 9) in Lynn, P., *Methodology of Longitudinal Surveys.* Sussex: John Wiley & Sons, Ltd.

Singh,R.(1983).On the use of incomplete farmes in sample surveys .Biom.J.25,545-549.

Singh,R.(1986).Predecessor-Successor Method. Encyclopedia of Statistical Sciences. V.7,137-139.John Wiley& Sons Inc.

Smith, T. W. (1984)“Estimating non-response bias with temporary refusals”, *Sociological Perspectives* , Vol. 27, pp. 473-489.

Steeth, C. (1981) "Trends in Non-response Rates 1952-1979", *Public Opinion Quarterly*, Vol. 45, pp. 40-57.

Stoop, L. (2005) "The Hunt for the Last Respondent", *The Hague, Netherlands: Social and Cultural Planning Office.*

Stoop, I., Billiet, J., Koch, A., and Fitzgerald, R. (2010) "Improving Survey Response: Lessons Learned from the European Social Survey", Chichester: John Wiley & Sons, Ltd.

Summerfield, M. (ed.). (2010) "HILDA User Manual- Release 9", Melbourne Institute of Applied Economics and Social Research, University of Melbourne.

Uhrig, N. (2008)"The Nature and Causes of Attrition in the British Household Panel Survey", *Institute for Social and Economic Research Working Papers,* No. 2008-05. Colchester: University of Essex.

Voogt, R. (2004): \I am not Interested: Non-response Bias, Response Bias and Stimulus Effects in Election Research," Ph.D. thesis, University of Amsterdam.

Watson, D. (2003) "Sample Attrition between Waves 1 and 5 in the European Community Household Panel", *European Sociological Review*, vol. 19:4, pp.361-378.

Watson, N., and Wooden, M. (2009) "Identifying Factors Affecting Longitudinal Survey Response", (chapter 10) in Lynn, P., *Methodology of Longitudinal Surveys.* Sussex: John Wiley & Sons, Ltd.