ISER Working Paper Series

Weighting for Non-monotonic Response Pattern in Longitudinal Surveys

Husam Sadig
Institute for Social and Economic Research
University of Essex

No. 2014-34
October 2014

INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

www.iser.essex.ac.uk

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Non-technical summary

Longitudinal surveys collect data from the sampled units on multiple occasions. At some data collection points, some sample members might, for any reason, not participate. This, leads to reduction in the data that were initially sought by the survey researcher. Also, if those who do not participate have different answers to the survey questions than those who participate, survey results that are only based on answers provided by those who participate, may be misleading. In any case, survey results might not be accurate. Thus, people who analyse the data sometimes try to compensate for non-participation by allowing the data provided by some participants, who are in some sense 'similar' to the non-participants, to have greater influence on the results. Survey organisations try to make this simple for data users by calculating the measures of the influence that each participant should have. These measures are referred to as 'weights'. The weights are thought to reduce the potential error due to non-participation by increasing the influence of participants who appear to be similar to those who did not participate.

However, the existing approach of preparing the weights in longitudinal surveys produces a single set of weights. This set of weights is designed for analyses of data from all data collection points. Nonetheless, since just one set of weights is available, analyses of data from sub-sets of data collection points will be forced to use it too. However, in the latter case, the potential error in the results might not be reduced as anticipated since the weights were not created specifically for the analysis of data from a sub-set of data collection points. Therefore, the existing approach may benefit from examining alternatives.

This research tests an alternative approach of preparing additional sets of weights, created to compensate for the data that are missing due to sample members not participating in specific combinations of data collection points. A relevant combination of data collection points may consist of ones where survey questions were asked on the same subject. I find that this approach to preparing weights is more successful in reducing the error in survey results than the single set of weights approach.

Weighting for Non-monotonic Response Pattern in Longitudinal Surveys

Husam Sadig

Institute for Social and Economic Research, University of Essex

Abstract

In longitudinal studies, analysis can be based on any one of a large number of wave-combinations. However, only one set of non-response weights (often based on respondents from all waves up to the latest) is typically offered on public use data files. We refer to this as a single weighting strategy (SWS). This paper uses data from the British Household Panel Survey to illustrate the limitations of the SWS. We evaluate the effect of designing weights based on response to wave-combinations concerned with the same module of questions. The analysis shows that the use of SWS may lead to an unnecessary loss of respondents if used with a different combination of waves. This leads to less precision on some, but not all, of the survey estimates.

Keywords

Non-response error, bias, precision, weighting, attrition.

Author correspondence address: hesadi@essex.ac.uk; husam.sadig@hotmail.com.

Acknowledgements: Thanks are due to Peter Lynn and Olena Kaminska for their constructive comments and helpful advice.

1. Introduction

Regardless of their aim, objective and design, surveys are usually afflicted by a constant set of problems such as non-coverage and non-response. In all surveys, some of these problems, particularly non-response, are unavoidable. Weighting strategies based on auxiliary information available for both respondents and non-respondents are used to resolve this problem by compensating for the missing information. In longitudinal surveys, however, in addition to successful use of powerful auxiliary information, developing non-response weights requires paying attention to other complex aspects of the survey. For example, for an accurate survey estimate, weights should be constructed using the set of respondents which will contribute to the estimation process. In its place, weighting based on some, but not all of the respondents who will be used to construct the estimate, may not tackle non-response bias adequately. Moreover, it is an inefficient use of data in the estimation since some data are ignored even though they contain useful information.

After many waves are conducted, data can be drawn for analysis from a number of possible combinations of waves. Different combinations are relevant to different analysis objectives. For every respondent, each combination of waves leads potentially to a different responding status (either responds or does not respond in all waves in the given wave-combination). As a result, each wave combination may result in a different set of respondents both in terms of number and composition. It is very likely that the number of responding units in larger combinations (wave-combinations that contain large number of waves) is smaller than the number of the responding units in wave-combinations with smaller number of waves. However, from a weight construction's perspective these facts are ignored. Usually, a single set of weights is constructed based on information about the responding status in one wave combination (typically all the conducted waves). Now, if the constructed set of weights is implemented in the analysis of a sample that is drawn from another wave-combination where the responding status – for some respondents- differ from the responding status in the wave-combination that is used to derive the weights, those who had a different responding status in the wave-combination under investigation, will be modified incorrectly. With respect to a given wave-combination, weighting will be more appropriate if it is carried out based on information on the responding status in this wave-combination irrespective to responding status in other wave-combinations. However, providing weight adjustments for all possible combinations of waves may be impractical, especially after a large number of waves are conducted. Nonetheless, it might be useful to provide weight adjustments to some combinations of waves. A challenge for the survey organisation then is to identify the best possible wave-combinations in order to specifically design their corresponding sets of weights. One way of doing this is to consider wave-combinations that obtain the same theme of data. Such combinations are likely to be in demand from analysts as a base for analysis.

This paper investigates whether the use of a single set of weight adjustments, which is created, based on one wave-combination is adequate to handle non-response error in data drawn from other wave-combinations. It also evaluates the choice of providing weights tailored to a combination of waves that carry the same module of questions. The paper reviews the principal aspects of non-response and weighting in survey research in general, but it mainly focuses on weighting for non-response. It begins with non-response problem and its weight adjustments in a cross-sectional design before transferring the concepts into a longitudinal context.

Although weighting incorporates a number of stages in order to compensate for all the missing units in the sample, due to different reasons for being missing either at the individual or a household level, this paper is only concerned with weighting to compensate for unit non-response at the individual level. Thus, the terms ‘weights’ and ‘non-response’ will be used in this paper to refer to non-response weights and unit non-response respectively.

2. Weighting survey data

After the data collection stage is completed, a weighting process is usually undertaken in order to produce valid results and reduce bias in estimates. Weighting is a process to assign more value to some of the eligible survey respondents so as to modify them to account for the sample design (unequal-chance of selection) and to represent individuals missing due to incomplete frame or non-response (Bethlehem, 2009; Biemer and Christ, 2008). Thus, ignoring the weights and naively treating the data as if they were selected with equal probabilities from a frame that covers the whole target population, with no non-response, often leads to biased estimates. On the other hand, allocating appropriate weights to respondents reduces the bias and enhances the reliability of the results by taking into account units that are not there because of coverage or non-response error. Thus, weighting is essential and highly recommended by statisticians to improve the precision of survey estimates. The weight is interpreted as the number of units in the population that are represented by the sample respondent (Särndal and Lundstrom, 2005; Lynn, 2005; Biemer and Christ, 2008).

Weighting is often used for one of three principal purposes: to balance for unequal probabilities of selection, to correct for non-coverage and/or to compensate for non-response.

Unequal probabilities of selection (sometimes referred to as the inclusion probabilities): these occur when some of the units in the population have a different chance of being in the selected sample than other units. If a sample has been selected with unequal probabilities, estimates such as the sample mean are biased. For example, consider a sample design that aims at randomly selecting one adult from each of 10 households that were randomly selected from 400 households. In this example the chance of an adult being selected from a certain household 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 families with fewer adults. This is the case even if the survey is deliberately designed to select with unequal probabilities. Sample weights are thus used to balance the unequal probabilities of selection and eliminate its bias.

Non-coverage (or frame error): this occurs when some of the units in the target population are not included in the sampling frame. Therefore, these units have no chance of being in the selected sample (Dillman, 2007; Massey, 1988). In fact, it might be impossible sometimes to create the perfect frame especially if the target population is very large. For example, using the telephone directories to list all the households that have landlines is inadequate since some numbers are not listed. Therefore, estimates from a sample that has been selected from this frame are biased unless a correction is made through statistical weights to compensate for the missing units. Non-coverage error is corrected through a weighting technique called 'post-stratification'. In post-stratification, respondents' weights are adjusted further so that the sums of the adjusted weights are equal to known population totals for certain sub-groups of the population. For example, if the population totals of subgroups defined by age groups and gender are known (maybe from the sampling frame or other external source), post-stratification adjusts the sample weights so that their distribution by the defined subgroups in the sample is the same as the known population distribution. However, non-coverage bias may not be eliminated if post-stratification is implemented: but it will be reduced (Biemer and Christ, 2008).

In this paper, I concentrate on the use of weights to compensate for non-response, as the other reasons for weighting are beyond the objectives of this paper.

3. Non-response

The term 'non-response' is used to describe the situation when the survey organisation is unable to collect responses from all of the sampled units. Non-response occurs in every survey as it is very difficult to obtain all the required information from the selected sample (Lynn, 2008; Särndal and Lundstrom, 2005). This includes even the most well designed surveys conducted by highly experienced survey organisations. Non-response can take one of two forms: one is unit non-response, when a sample member does not provide (for a variety of reasons) response for the survey. The second, item non-response, occurs when a sample member responds to some of the survey questions but fails to provide answers for a particular item (Madow et al, 1983; Lessler and Kalsbeek, 1992). This paper is concerned with unit non-response.

3.1 Reasons for unit non-response

Non-response arises for a wide range of reasons. The literature on non-response generally distinguishes among five different reasons for non-response (see Lynn, 2008; Groves et al, 2004; Kish, 1965): (1) *failure to locate sample unit*: it is sometimes impossible for the survey researcher to successfully locate sampled members if, for example, their address is incomplete, (2) n*on-contact:* this refers to a situation where a sampled member is successfully located but the interviewer is unable reach them to conduct the interview because the sampled member is not available (e.g. not at home). This depends to an extent on the survey population; for instance, in eastern countries it is more common to find a woman than a man at home in the daytime (Kish, 1965), (3) *refusal:* this happens when a successfully located and contacted sampled member explicitly refuses to take part in the survey. This depends mainly on factors such as the nature of respondents; some individuals are more cooperative than others due to culture, social classes and the demographic categories they belong to. Also, the survey objective and the skills of the interviewer are important factors affecting sample members to respond to the survey, (4) i*ncapacity or inability:* in this case, the approached sampled member may be willing to take part in the survey; however, he or she is unable to due to illness, illiteracy or language barrier, and (5) *loss of information:* this refers to the accidental loss of data after the field work. For example, questionnaire forms might get lost in the post or destroyed if neglected.

Furthermore, the reason for non-response may also be linked to one of four issues: survey topic, type of sample units, survey design and the mode of data collection (Lynn, 2008). For instance, surveys with sensitive topics (e.g. self opinion about same sex marriage) are expected to suffer from high non-response rates. Besides, surveys that target households as sample units usually achieve higher response rates compared to surveys that collect data from individuals, since the possibility of successfully contacting the unit in the former is higher. Also, there is evidence in the literature that face-to-face interview surveys achieve higher response rates than telephone and mail surveys (Groves et al, 2004).

3.2 Effect of non-response

Non-response leads to one of two problems[1]: (a) if many sample members do not participate in the survey, the sample size that one had hoped for at the design stage will be reduced. Thus, estimates derived from the smaller sample would have bigger standard errors and hence less precision. This is however a minor problem, as the sample size can be set to a required achieved sample according to a predicted non-response level (Lynn, 1996); (b) if many sample members do not respond to the survey, and those who do not respond are different in their characteristics from respondents in terms of the survey key variables, estimates based solely on information from respondents can be biased (Lessler and Kalsbeek, 1992; Särndal and Lundstrom, 2005).

3.3 Link between Response rate and non-response bias

In many surveys, one of the main concerns is to increase the response rate. Increasing the response rate is desirable since it automatically decreases non-response rate and hence may minimize the likelihood of the bias linked to non-response. This indicates that non-response bias strikes surveys with low response rate more than surveys with high response rate. However, the magnitude of the response rate does not provide information about the existence or the size of non-response bias (Groves, 2006; Groves and Peycheva, 2008). For example, having a low response rate does not necessarily lead to a high non-response bias as the bias only exists if there is a difference between respondents and non-respondents in characteristics related to the survey key variables. In fact, with a low response rate non-response bias might not even exist if respondents and non-respondents are very similar in all the characteristics related to all the survey key variables. This implies that the survey organisation has to examine the existence of non-response bias before deciding to deal with it. In other words, survey researchers have to detect if there is a difference between respondents and non-respondents before designing weights to reduce or remove non-response bias. Nonetheless, it is impossible to discover differences between respondents and non-respondents using data with respect to the survey key variables, because it is only available for respondents (Bethlehem, 2009). However,

[1] In the same survey, the effect of non-response varies across estimates. For example, if respondents are systematically different from non-respondents on a variable ‘y’ but similar to non-respondents on another variable ‘x’, estimates derived from ‘x’ will be less affected by non-response.

auxiliary data that are available for both respondents and non-respondents can be used to inspect the differences between respondents and non-respondents before dealing with the problem.[2] Therefore, in recent years, the link between response rate and non-response bias has been a hot topic among survey researchers. Although many survey researchers stress the importance of endeavouring to increase the response rate (see for example Alreck and Settle, 1995), several studies on the other hand found that changes in response rates may not necessarily have an impact on the survey estimates (see Curtin, Presser and Singer, 2000; Merkle and Edelman, 2002).

There are generally two procedures for dealing with non-response problem. One is to minimise non-response when collecting the data (Lynn, 1996; Stoop et al, 2010). Another approach is to deal with non-response at the analysis stage. One way of doing this is to incorporate weight adjustments before analysing the data to compensate for non-respondents. Although the former could be very useful in maximizing the response rate, it is impossible to obtain 100% response rate, especially in surveys targeting households and individuals. Therefore, weighting may be needed to compensate for the remaining lost sample units. Nevertheless, most of the available literature encourages survey organisations to combine both methods in order to tackle non-response effectively. In this paper, the discussion is limited to the weighting approach.

4. Constructing non-response weights

The most desirable feature of sample surveys is that they enable survey researchers to make inference about a large population using a manageable segment of that population. However, this assumes that the sample is representative of the population. This assumption is satisfied if sample units are selected through a probability sample design (randomly) and that every unit in the sample provides data that are usable to make the inference. Nevertheless, with non-response, this assumption is violated since there are no data available for some of the sample members and, if the sample is used as it is, it may not be representative of the population.

[2] The literature on non-response bias suggests approaches to examine the existence of non-response bias. For example, Groves (2006) provides methods for assessing non-response bias and reviews their strengths and their weaknesses.

For example, suppose that a survey is conducted to estimate the proportion of single-person households in a town. There is evidence in the literature that single-person households are difficult to contact since they are less likely to be available (see for example Uhrig, 2008). Thus, the responding sample may only include a small number of single-person households. This will lead to underestimating of the proportion of single-person households. Therefore, unless a correction is made to compensate for the non-responding single-person households, the survey estimate (proportion of single-person households in this case or any estimate related to it) will be biased.

Non-response weighting is a correction that is implemented specifically to compensate for the units missing from the sample. It adjusts the responding sample so that its distribution is the same as the selected sample, and hence produces more accurate and unbiased estimates.

Although the rationale behind weights is convincing and well established, there is no constant universally held protocol to compute them. Weight 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. However, there are general well-known steps to constructing the weights, to compensate for non-response, which should be followed in order to get a higher quality set of weights.

The final weight to compensate for non-response (w_i) consists of two weight components. The first component is referred to as the design weight. The design weight represents the starting point for constructing the final non-response weight and it is derived from the sample design.

Design Weight

The design weight (or the base weight) is developed to account for the unequal probabilities of selection of the sample. If the sample is selected with unequal probabilities of selection, sample units do not represent the same number of units in the population. The design weight for the i^{th} unit in the sample (w_{Di}) reflects the number of cases in the population that are represented by this unit from the procedure of selection perspective. For instance, consider the same example in section (2) about selecting one individual from each of 10 households that were randomly selected from 400 households. The probability of selecting an individual differs across the 10 households depending on the size of the household. As a result, individuals from smaller size households will have higher chances of being included in the sample. Thus, the survey results can be biased towards small-size households unless an adjustment is implemented to balance the probability of selection. This adjustment 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, simply, the design weight is the inverse of the selection probability (p_i) for each unit in the sample.

$$w_{Di} = 1/p_i \qquad (1)$$

If sample units were selected using a simple random sampling method, p_i becomes constant. Thus, 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 due to the different probabilities of selection.

The second component of the final weight to compensate for non-response is non-response weight (w_{NRi}). It is used to account for the missing units in the sample.

Non-response weight

The probability of responding to a survey differs across people of different characteristics. Some individuals have higher propensity to respond than others. Non-response weight is 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).

There are two ways to estimate the response propensity for each unit in the sample in order to calculate non-response weight: weighting classes and estimation through a binary outcome regression model.

Weighting classes

Weighting classes is a simple approach that involves dividing the sample into h sub-groups using the main variables of the survey that are known for both respondents and non-respondents. These sub-groups are referred to as weighting classes. The response propensity (R_h) is then calculated for each weighting class by dividing the number of responding cases in that class by the total number of cases (i.e. the class response rate). The non-response weight for the ith sample unit in class h is then given by:

$$w_{NRhi} = 1/R_h \qquad (2)$$

The disadvantage of this method is that classes are subjectively identified in one or two dimensions, by using one or two key variables to produce a range of response propensities. Also, classes with low response propensity attract large weights and consequently introduce large variance. Lynn (1996) suggests avoiding weighting classes with a response propensity that is less than one-fifth of the overall survey response rate.

Binary outcome regression model

Using a binary outcome model to estimate the response propensities for each respondent in the sample is more effective (Grau, 2006). In this method, a binary outcome model, for example, logistic regression is used usually to predict the probability of response, based on variables that are correlated with the response propensity and available for respondents and non-respondents. The non-response weights can then be calculated as the inverse of the predicted values from the model.

Consider R_i to represent the dependent variable in the logistic regression. R_i can take the following values:

R_i= 1, if the ith sample unit responds.

R_i= 0, otherwise.

The independent variables (X_{ij}) are a number of variables available for both respondents and non-respondents and are thought to be related with R_i.

In this context, the logistic regression is used to estimate R_i. Non-response weights are then computed as:

$$w_{NRi} = r_i^{-1} \quad (3)$$

Where r_i^{-1} is the inverse of the predicted value of R_i.

The advantage of using a model to estimate R_i is that dummy and continuous variables can be combined to fit a range of models, and therefore obtain more effective non-response adjustments (Biemer and Christ, 2008). However, an important disadvantage is that the instability of w_{NRi} could lead to large variation in weights. Nonetheless, the estimated response probabilities are sometimes grouped into weighting classes, and weights are recalculated using either the mean predicted probability in the class or the observed response rate in the class in order to reduce the variation among individual response propensities.

Since being selected in the sample is independent of responding to the survey, a final weight is constructed as the product of the design weight and non-response weight. This way, every unit in the sample is adjusted using its level of chance of being selected in the sample and its tendency to respond to the survey simultaneously.

The final analysis weight to correct for non-response (w_i) for the ith case in the sample is:

$$w_i = w_{Di} * w_{NRi} \quad (4)$$

Where w_{Di} is the design weight and w_{NRi} is the non-response weight for case i.

5. Non-response in Longitudinal Surveys

Non-response and the notion of weighting that are introduced in sections (3) through to (4) assume a cross-sectional survey design. In this section these concepts are expanded into a longitudinal context.

A distinctive feature of longitudinal surveys is that they collect observations from individuals on multiple occasions (Lynn, 2009). This design involves following individuals over time and continuing to collect data from them. However, respondents might not be available to participate in the survey every time data are collected. Therefore, non-response can occur for a number of reasons that result from non-contact or refusal (Lepkowski & Couper, 2002). For example, unreported change in the residence address might result in non-response; also, some respondents may refuse to respond at some point although they have participated in previous waves. Thus, the complexity of longitudinal surveys turns non-response into a dynamic event that accumulates over time when further waves are conducted (Watson and Wooden, 2009). This may be a dilemma for the survey organisation, especially if respondents were chosen via a probability sample design since they cannot be replaced.

Non-response in longitudinal surveys can be in one of two forms: (1) *Wave non-response* refers to the process where a respondent is absent from the survey for at least one wave but returns to the survey in a later wave. (2) *Attrition* on the other hand occurs when a respondent participates in the survey for one or more waves but permanently stops participating at some point during the survey course. Although the former is not trivial, survey researchers are more concerned about the latter for at least three reasons: (a) more information are lost in the case of attrition; (b) the potential bias caused by non-response is more likely to occur; (c) any observed information collected in earlier waves become weak predictors as more waves are added (Chang, 2010). In surveys with an indefinite number of waves, it is always difficult to distinguish wave non-response from attrition as it is controlled by the respondents' behaviour in the future. In contrast, the point of attrition can be identified in finite length surveys when no further waves are conducted (Uhrig, 2008).

5.1 Causes of Attrition and Wave Non-response

Attrition and wave non-response are special cases of non-response occurring in panel studies. The causes of non-response in longitudinal surveys may be similar to those in cross-sectional surveys but somewhat different in at least two ways:

(a) In a longitudinal survey respondents are burdened by a constant long-term commitment to responding.
(b) Changes occurring during gaps of data collection points may have a big effect on the response process.

Therefore, some reasons for non-response may be specific to longitudinal studies. For example:

(1) *Failure to update contact information:* If survey participants move houses between waves without informing the survey organisation, it could be too expensive if not impossible to track them and failure to do so will directly result in non-contact.

(2) *Loss of interest:* Although some respondents may be interested in taking part in the survey at the start, their level of willingness to continue giving data at every wave is a function of the survey organisation's effort to maintain their interest level (e.g. use of incentive). Failure to retain participation interest results in refusal.

(3) *Changes in health condition*: Longitudinal surveys are conducted over a long period of time. Over this period, some respondents might suffer from a bad health condition leading them to have to drop out.

(4) *Technical issues related to data collection strategy*: Survey organisations may be forced sometimes to adopt changes in the data collection strategy. For example, a failure in maintaining the interviewer from the last wave may affect the propensity to respond in the current wave. This is particularly common when the new interviewer is less experienced. Although there is no evidence that the change of interviewer results in non-response, recent research showed that interviewer continuity is associated with low propensity for refusal (Lynn, Kaminska and Goldstein, 2014).

Although both attrition and wave non-response are considered as forms of non-response in longitudinal surveys, the causes for each may be different. Still, neither of the two types is desired in longitudinal surveys.

5.2 The effect of non-response in the longitudinal context

During its course, every longitudinal survey must experience respondents dropping out at some point and may or may not participate at a later data collection point. Regardless of the reasons leading to losing survey participants during the study, in panel studies, non-response is problematic for at least three reasons:

(a) the original sample may suffer from a monotonic decrease in its size. As a result, after a large number of waves, the survey organisation might end up with a relatively small sample that is incapable of producing precise estimates. In section (3) it is mentioned that this is not a major problem in cross-sectional surveys since a required achieved sample size can be set according to a predicted non-response level (Lynn, 1996). In contrast, in longitudinal surveys, even if the survey organisation invested in a very large sample size (as is the case in Understanding Society,[3] the world's largest longitudinal survey) the reduction of the sample size may still be a problem in the long term. In other words, whilst obtaining a large sample size may be the cure for one of the effects of non-response in cross-sectional surveys, it would be a delaying factor in longitudinal surveys, particularly in panels of indefinite length. This distinction shows that the decrease in the sample size due to non-response is more problematic in longitudinal surveys, and in fact it is a typical feature of panel data. Therefore, it would be very wise for survey organisations to establish future plans at the design stage to deal with this problem in order to increase the size of the achieved sample and hence ameliorate the precision of the survey estimates.

(b) the process in which respondents drop out of the study may not be random (Fitzgerald, Gottschalk and Moffitt, 1998; Watson, 2003) and it is plausible to assume that some of the drop outs are because of the survey topic. For example, respondents may participate in one or two waves and then decide to drop out after they have discovered what the survey is about. If these drop outs are different from respondents in terms of what the survey is measuring, the sample becomes progressively unrepresentative of the

[3] Understanding Society is a large survey conducted in the UK since 2009 to provide evidence of individuals' lives, behaviours, experiences and beliefs. With a sample of 100,000 individuals, Understanding Society will take over as the world's largest survey of its type.

population as more waves are conducted. Consequently, estimates derived from the achieved sample will be biased.

(c) unlike non-response in cross-sectional surveys which occurs at a single point in time, in longitudinal surveys, non-response can follow a pattern over time. This pattern depends on the policy of data collection implemented in the survey[4]. For example, considering the British household Panel Survey (BHPS) which attempts to collect data from sample members at every wave regardless of their previous responding statuses, table 1.0 shows that, with just three waves, there are seven possible combinations of waves in which a sample member might respond. Also, there would be seven combinations of waves from which data could be used for analysis. However, each of these seven patterns (or wave combinations) may have a different sample size (inconsistency). In general, responding in a larger number of waves is associated with a smaller sample size and vice versa. A side from producing different precision levels of the same estimate calculated from different wave-combinations, this inconsistency in sample sizes may lead to complications on other aspects of the survey such as weighting as it is explained in the next section.

[4] There are many data collection policies. The major of these: attempt collecting data from eligible cases at every wave regardless of whether they participated in a previous wave or not, attempt collecting data only from wave 1 responding cases or attempt collecting data only from cases responding to previous wave. In this paper, we assume a policy where data collection attempts are conducted at every wave regardless of participation in previous waves.

Table 1.0 Number of respondents responding in different wave-combinations in the first 3 waves of the BHPS

	Wave 1	Wave 2	Wave 3	Number of respondents	Percentage of respondents
1				10248	89.18%
2				9845	85.68%
3				9600	83.44%
4				8970	78.06%
5				8736	76.02%
6				8419	73.27%
7				8170	71.10%

* The shaded areas indicate 7 possible wave-combinations from which analysts might use data for their analysis. The numbers and percentages reflect the size of the sample that will be obtained in the relevant wave-combination**.**

As is the case in cross-sectional surveys, the non-response problem in longitudinal surveys can be dealt with either at the data collection stage or at the analysis stage by implementing a form of adjustment to compensate for non-respondents. This paper focuses on the use of weight adjustments in dealing with non-response error.

6. Non-response weights in longitudinal surveys

In longitudinal surveys, a central role for the survey organisation is to prepare weights to adjust for non-response and include them in public use data files (Lynn and Kaminska, 2010). Designing non-response weights from cross-sectional data is straightforward in comparison with longitudinal data. This is because in cross-sectional surveys observations are recorded at a single point in time so that the response process can be defined by a binary variable (non-response=0 and response=1) reflecting the response status in the single time point. Moreover, it is known that the developed weights will compensate for non-respondents at the defined single point in time when the data were collected. In contrast, in longitudinal studies data are collected at multiple occasions, so respondents have many records in the data set, each referring to a different data collection point. Thus, the variable representing the response process can have a large number of categories, each category identifies response outcome in a certain combination of data collection points. This complexity implies changing the weights as time goes on. Furthermore, it permits designing different sets of non-response weights from different combinations of subsets of data. Moreover, if a set of weights is derived from a particular subset of data which is linked to certain data collection times, it cannot be asserted that it can compensate for non-respondents with respect to a different subset of data that is connected to different data collection times. This is because the set of respondents in the two subsets of data can be different. For example, in a survey with a limited number of waves, say 15, assume that the set of non-response weights at wave 15 is designed by reference to responding in all the waves together up to wave 15. This set of weights can be used to correct for non-response in any analysis that requires a sample of respondents who responded at all of the 15 waves. However, supposing that the analysis was to be carried out using data only from the last five waves, the sample used in the analysis in this case may be different than the sample used for the analysis on data from all 15 waves (likely to be larger as it consists of sample members who responded in wave 11, 12, 13, 14 and 15 regardless of whether they also responded at any other wave). Thus, the set of non-response weights at wave 15 would then be suboptimal in this case, since it rules out all sample members who did not respond in all of the 15 waves by assigning a weight of zero to them, even if they have responded in the required last five waves. Clearly, this is an unnecessary loss of information which could be avoided if a set of weights was designed specifically for respondents in wave 11, 12, 13, 14 and 15.

Therefore, there are two dimensions need to be thought of when creating non-response weights for longitudinal data. First, non-response weights are not fixed over time i.e. any set of weights that is created at a certain wave needs to be updated when the next wave is conducted as the set of responding units will be updated as well. Second, the multi-wave feature in longitudinal surveys allows for data to be drawn for analysis from different combinations of waves. However, the set of responding units can differ across wave-combinations offering potentially different subsamples for every possible combination. Thus, weights may be required for a number of combinations of waves too, as one set of weights might not be sufficient in handling non-response error in data from all wave-combinations.

However, in the major longitudinal surveys in the world, weighting for non-response is often a single weighting strategy overlooking the fact that different wave-combinations can potentially provide different sets of respondents. For instance, in the British

Household Panel Survey (BHPS), longitudinal weights at any wave 'w' are only available for a balanced panel from all waves up to wave 'w' (Taylor *et al*, 2010). Likewise, longitudinal weights in a current wave in the Swiss Household Panel (SHP) are designed to extrapolate to the population living in Switzerland at that wave using respondents from all waves up to the current (Plaza and Graf, 2008). This is also the case in the German Socio Economic Panel (GSOEP) and the Panel Study of Income Dynamics (PSID), where no particular combination of waves are provided with specially designed longitudinal weights; instead, weights in the latest wave are available for the set of respondents from all waves including the latest (Kroh, 2009; Gouskova, 2001). However, in the Household, Income and Labour Dynamics in Australia (HILDA) a different procedure is considered where weights are also provided for a balanced panel from every combination of two waves (Summerfield, 2010).

This single non-response weighting strategy, which is used in almost every survey, could be helpful and practical in reducing non-response bias, but may be inadequate in respect to the subsample being used for analysis.

In theory, the way out of this problem is to design a subset of non-response weights for every possible combination of waves. However, providing weights for all possible combinations of waves might not be achievable in practice sometimes. For example, after k waves are conducted, there is a (2^k-1) possible combination of waves to provide weights for. Thus, this number increases rapidly when more waves are added, and it could even outnumber the number of variables in the survey in a long term panel. Moreover, in practice, not every possible combination of waves is of use for data users. Therefore, another way of dealing with this problem is to produce a limited number of subsets of weights for a limited number of wave-combinations. The choice of these wave-combinations can be guided by the interest of data users to analyse data from the given wave-combinations.

Thus, as shown in figure 1, there are at least three possible strategies of weighting: single set of weights, all possible sets of weights and limited number of sets of weights. The limitations of the first and the second strategies are that they could be suboptimal and impractical respectively. In contrast, the advantage of the third strategy is that it is at least practical. Nevertheless, it is a challenging task to identify combinations of waves that will be of interest for data users. But the possibility that a single weighting strategy might not be sufficient generates interest in the development of more subsets of weights. Hence, the investigation of this is an important aspect of weighting panel data.

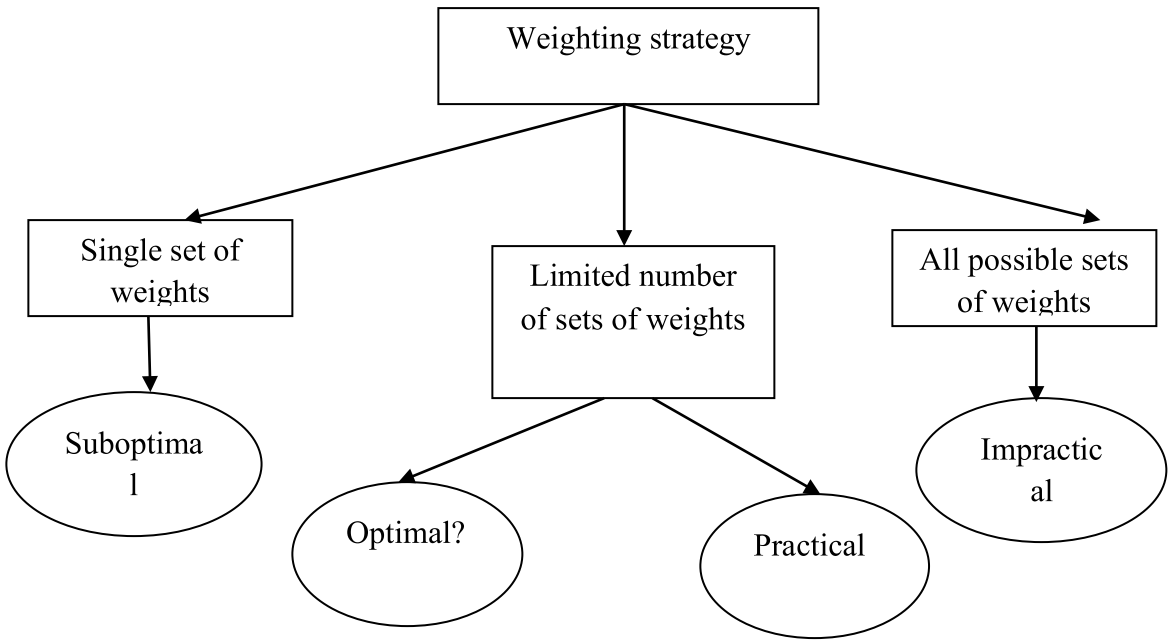

Figure 1. Three types of weighting strategy.

Very little work has been done in this area. In fact, the only effort I have come across is by Lynn and Kaminska (2010), suggesting criteria for developing subsets of longitudinal non-response weights. According to Lynn and Kaminska (2010), the following criteria should be considered when choosing wave-combinations, in order to produce the best subsets of weights:

Survey Design: If a certain combination of waves is not available in the survey by design (for any reason), it can be excluded. They give an example by stating "*if a survey has a rule not to attempt data collection from any unit that has been non-respondent in three consecutive waves, then all combinations involving a respondent wave following three or more non-respondent waves can be dropped*".

Analytic Use: Weights should be produced for combinations that are more likely to be wanted by analysts.

Level of Non-response: Consider Ω to be a set of a number of wave-combinations. If the sample size in all these combinations differs only by one or two cases, there is no need to provide weights for every combination in Ω as it is unlikely to make much difference. Hence, one subset of weights could be defined for all the combinations in Ω.

Correlates of Non-response: If the attrition pattern is very similar amongst a consecutive set of waves, weights designed for a subset of waves might be similar to weights designed particularly for the whole set. This does not include wave 2 and 3 as the attrition in them is believed to be distinctive.

Impact on Estimates: This could be used to judge the consideration of the criteria used to identify the combination of waves for which weights should be produced. For example, subsets of weights that produce the same estimate as others could be dropped.

Although the idea of producing subsets of longitudinal weights might seem adequate, the complexity of longitudinal surveys makes it difficult to establish a single choice method to decide upon the necessary subsets of weights. Nonetheless, indicating the state of current knowledge, the notion of subsets of weights has not yet been applied in the major longitudinal surveys in the world.

A common feature of longitudinal surveys is a frequently asked module of questions where certain waves are conducted to obtain information about specific topic(s). For example, wave 8, 13 and 18 in the BHPS provide data on neighbourhood, expectations of relationships and marriage in future. Also, waves 5, 10 and 15 are concerned with data about wealth, assets and debt. Thus, it might be useful to provide BHPS data users with a subset of weights designed specifically for the analysis of data from these waves.

In this paper, data from BHPS are used to examine the difference (in terms of impact on estimates) between a set of weights that is designed specifically for waves 5, 10 and 15 and the weight adjustments that are normally used (longitudinal weights at wave 15) in the context of a question module that is asked in waves 5, 10 and 15. Section 8 provides a detailed description of how the investigation is conducted. Conducting this investigation will reveal whether the single weighting strategy is inadequate in handling non-response error in a subset of waves (5, 10 and 15). Furthermore, it will evaluate the choice of producing subsets of weights for wave-combinations that are concerned with a rotating module of questions.

7. Non-response weighting variables

The choice of variables used to create non-response weights (sometimes called auxiliary variables) plays an important role in removing non-response bias. In order for non-response weights to be effective in removing bias, auxiliary variables have to be highly correlated with the survey key variables and the response propensity (Bethlehem, 2009; Groves, 2006). This might sometimes require the survey organisation to obtain such variables at the data collection stage, as the way to successful non-response weights depends largely on the choice of powerful auxiliary data (Särndal and Lundstrom, 2005). In recent years, survey researchers have laid the foundation for principles to guide the selection of the best set of auxiliary variables to correct for non-response. In general, a variable is said to be powerful in removing the 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 (Särndal and Lundstrom, 2005).

In order to model the response process and create weights, it is necessary to use variables that are available for respondents and non-respondents. Weighting variables can be drawn from many sources. These sources could be from inside or outside 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 (paradata); for example, what was the mode of data collection (e.g. phone, web, mail, or in person). (b) variables based on the interviewer's observations about some characteristics related to the household and/or individual (e.g. type of accomodation). (c) variables taken from the sampling frame; these are usually available if the sample is taken from administrative records (e.g. levels of proficiency or educations). (d) variables linked from another data base; 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), although the postcode itself does not provide information about respondents living at the selected address, it can be used to link geographical information from another data base such as credit scores (Lynn, 1996). (e) substantive survey variables available in previous waves (in the case of a longitudinal survey).

While longitudinal surveys are fortunate with (e), some of the available literature focus on (a), (b), (c) and (d) (see Plewis, 2011; Kreuter and Kohler, 2009; Lynn, 2003; Groves and Couper, 1998; Lynn, 1996). The advantage of paradata, interviewer observations-driven variables, sampling frame variables and variables used to link information from another data base is that they are cheap if not completely free and can be available for every unit in the sample. For instance, variables related to the accommodation type, neighbourhood characteristics, time interviewer arrived to the house, and number of previous contact attempts do not require respondents to report them; instead, they can be observed by the interviewer.

Cheap-information source variables are largely and successfully used in the literature to correct for non-response. For example, to predict survey participation Bates et al (2008) used data about contact attempts from the National Health Interview Survey. According to Kreuter and Kohler (2009), contact sequences should also be taken into account when weighting for non-response. Lynn (1996) showed how information about the level of qualification gained at school which was available in the sample frame was used to analyse the response rate in the Scottish School Leavers Survey (SSLS). Lynn (1996) demonstrated the way of which the post code (the sampling frame) 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 analysed the response rate accordingly.

In contrast, most research has found variables such as sex, race, age, socioeconomic status, income and level of education to be excellent predictors of the response propensity and hence powerful weighting variables. For example, 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) declares that in GSOEP data on household and interview characteristics measured in 2007 such as change in interview mode, gender, age, job status, income and savings were used to predict the probability of re-interviewing versus refusal in 2008. Using the PSID, Fitzgerald et al (1998) showed that the level of survey participation is low among lower socioeconomic status individuals, unstable income earners and people with unstable marriages or migration histories. However, they indicated that these variables explain very little of the attrition in the sample and hence are not significant. Nicoletti and Buck (2004) reported that even when contact is made, actual response is higher among women than men in the BHPS. Both Becketti et al (1988) and Fitzgerald et al (1998) show that, excluding young respondents, attrition is positively associated with old age. Investigating attrition in BHPS, Uhrig (2008) found that housing tenure, marital status, size of household, sex, 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 longitudinal surveys, regardless of the variables used to correct for non-response, the importance of designing more than one set of weights has been neglected. Weighting is ideal if it is designed through the best set of auxiliary variables and used for the analysis of data drawn from the wave-combination that provided information to design the weights. Instead, non-response weighting adjustment in almost every survey is a single approach that uses information from one wave-combination but provides weights for usage with a number of combinations of waves. Thus, analysts who would like to use weights in their analysis has no choice but to use the available set of weights regardless of the wave-combination that they are using as a base in their analysis.

8. Methodology

The data for this study are from the British Household Panel Survey (BHPS): specifically, data from wave 1 to 15 with a specific focus on the combination of waves 1, 5, 10 and 15.

The choice of wave-combination

According to Lynn (2006), in spite of the fact that BHPS aims at providing cross-sectional estimates of the population in the UK during the period of the survey, its main purpose is to explore the dynamics of change experienced by the population under study. Moreover, BHPS is principally conducted so that primary 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, 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.

Although all waves generally provide data to be used for analysis in many of the social science disciplines, some waves are designed to cover certain components extensively. For instance, wave 11 and 16 provide data on ageing, retirement, family support, health and quality of life whilst data about wealth, assets and debt is available in wave 5, 10 and 15.

Therefore, data from such subsets of waves might be required for analysis frequently. However, as explained in section 6, BHPS does not provide subsets of weights that are designed especially for the analysis of these combinations of waves. For example, if a researcher wanted to do analysis on wealth, assets and debt, which involves using data from wave 5, 10 and 15, the weights available for this would be the longitudinal weights based on the respondents at all waves up to wave 15.

This research used data from wave 1, 5, 10 and 15 from the BHPS and designed a subset of non-response weights for this combination. Also, another set of non-response weights is designed based on respondents at all waves up to wave 15. The two sets of weights were created using the same variables and the same method (see next section). Thus, any potential difference between the two sets of weights will be due to differences in the two wave-combinations in terms of the responding sample as other factors were held constant. Analysis was carried out on savings and debts data using the two sets of weights. The issue of interest here is to compare estimation results produced from the use of the two sets of weights and conclude on which set of weights is better based on these.

Construction of Longitudinal Weights

Both sets of weights were created using a model-based method. The analysis was restricted to respondents aged 16 or above and alive during the course of the 15 waves. For both sets of respondents, logistic regression was used to estimate the response propensity in each case. Two indicators R_{1i} and R_{2i} represented the independent variables in each model. R_{1i} and R_{2i} take the following values:

$$R_{1i} = \begin{cases} 1, & if\ unit\ i\ responded\ in\ wave\ 1, 5, 10\ and\ 15 \\ 0, & otherwise \end{cases} \tag{5}$$

$$R_{2i} = \begin{cases} 1, & if\ unit\ i\ responded\ in\ all\ waves\ from\ 1\ to\ 15 \\ 0, & otherwise \end{cases} \tag{6}$$

The relationship between the response propensity in our two wave-combinations and the weighting variables may vary from a wave to another and between wave-combinations too. This can be because the values of some variables can change overtime for the same respondent (time-varying variables) allowing for a different probability of responding. For example, considering the combination of waves 1, 5, 10 and 15, some respondents may be unemployed at wave 1 in which they are easier to make contact with. But, they may become employed at wave 15 which makes it more difficult to contact them and hence this may result in non-response. Thus, one way of creating weights for this wave-combination might involve modeling the response in wave 5 conditional on responding in wave 1, modeling the response in wave 10 conditional on responding in waves 5, and modeling the response in wave 15 conditional on responding in waves 10. Each model can be estimated using variables from the previous waves in that combination in order to take into account the potential effect of the time-varying variables on responding. An overall non-response weight can then be calculated by multiplying weights produced from the three models. Another way is to ignore the effect that time-varying variables may introduce in the process of responding and use variables from one wave (usually wave 1) to create the weights through a single model. The latter approach can produce a more parsimonious model which has the advantage of avoiding the risk of inflating the variance due to weighting.

In this paper, a large mixture of continuous and categorical variables from wave 1 was used to model the response propensity in the two wave-combinations. Namely, these variables are: sex, age, ethnic group, region, health status, household size, presence of children in household, housing tenure, income, number of people age 75+ in the household, type of household, number in employment in household, education, employment status, savings, debt, type of accommodation, financial situation, socioeconomic group, number of weekly working hours, number of weekly overtime hours, work location, smoking status, car ownership, number of own children in the household, presence of others during interview, interviewer sex and length of interview. These variables were chosen from three categories of variables that are thought to affect the response propensity. These are: interview and interviewer characteristics (e.g. interviewer's sex and length of interview), household characteristics (e.g. household size and household type) and individual characteristics (e.g. age, sex and savings).

A small number[5] of respondents joined the BHPS after the first wave; those have been excluded from the analysis as there are no available data for them at wave 1. Two logistic models were estimated to explain the variation in the response propensity in wave 5, 10 and 15 and in all the waves up to wave 15 conditional on responding at wave 1.

$$\text{Logit}\ (R_{1i}) = f\ (\textstyle\sum_k I_{1k} + \sum_{jk} H_{1jk} + \sum_{ijk} D_{1ijk} + \varepsilon_{1i}) \tag{7}$$

$$\text{Logit}\ (R_{2i}) = f\ (\textstyle\sum_k I_{2k} + \sum_{jk} H_{2jk} + \sum_{ijk} D_{2ijk} + \varepsilon_{2i}) \tag{8}$$

Where:

$R_{1i} \equiv$ Responding Status at wave 1, 5, 10 and 15.

$R_{2i} \equiv$ Responding Status at all waves up to wave 15.

$I_k \equiv$ Interview and Interviewer characteristics.

$H_{jk} \equiv$ Household characteristics.

$D_{ijk} \equiv$ Individual characteristics.

$\varepsilon_i \equiv$ Error term.

The non-response weights for the two sets of respondents were then calculated as the inverse of the predicted value from the fitted model as shown in equations (9) and (10).

$$w_{NR1i} = \begin{cases} \frac{1}{r_{1i}} & if\ unit\ i\ responded\ in\ wave\ 1, 5, 10\ and\ 15 \\ 0, & otherwise \end{cases} \tag{9}$$

[5] This is a small proportion of those who resulted in non-contact at wave 1, but they were contacted at wave 2.

$$w_{NR2i} = \begin{cases} \frac{1}{r_{2i}} & if\ unit\ i\ responded\ in\ all\ waves\ up\ to\ 15 \\ 0, & otherwise \end{cases} \tag{10}$$

Where:

w_{NR1i} ≡ Case i non-response weight based on respondents in wave 1, 5, 10 and 15.

w_{NR2i} ≡ Case i non-response weight based on respondents in all waves up wave 15.

r_{1i} ≡ Predicted value of response propensity from the first model.

r_{2i} ≡ Predicted value of response propensity from the second model.

In order to check if the two sets of weights lead to different results, w_{NR1i} and w_{NR2i} were incorporated in modelling the change in two variables: Savings and Debts from wave 5, 10 and 15. The process of modelling each variable is explained in the next section. However, before applying the non-response weights on the data, w_{NR1i} and w_{NR2i} were multiplied by wave 1 non-response weights. This set of weights was provided by BHPS. In BHPS wave 1 non-response weight is a product of two weight components. The first is a weight to adjust for the variation in the inclusion probabilities. The second weight component is to compensate for non-response at wave 1. Thus, any of the final two sets of weights corrects for the differences in the selection probabilities and non-response in wave 1 and non-response in its wave-combination simultaneously. The final two sets of weights can be written as:

$$w_{1i} = w_{Di} * w_{NR1i} \tag{11}$$

$$w_{2i} = w_{Di} * w_{NR2i} \tag{12}$$

Where:

w_{1i} ≡ Case i final weight based on respondents in wave 1, 5, 10 and 15.

w_{2i} ≡ Case i final weight based on respondents in all waves up to 15.

w_{Di} ≡ Case i wave 1 weight.

Modelling savings and debts using the longitudinal weights

The British Household Panel Survey provides detailed information on savings and debts at the individual level for the years 1995, 2000 and 2005, representing waves 5, 10 and 15 respectively. In each of these waves, respondents were asked if they have money in savings and whether they owe money. If respondents have money in savings and/or owe money, they are then asked to state the amount held in these. This setting permits two main sets of dependent variables which were used in the analysis: (a) Dichotomous: these are two variables, one indicates whether an individual has savings or not and the other indicates if the individual is in debt; (b) Continuous: these are two variables reflecting the amount of money held in savings and debts.

Across the three waves, the proportion of missing values among the first category of the dependent variables is negligible (less than 1%). However, the second category of the dependent variables shows a large number of missing values[6] among eligible respondents across the three waves. Therefore, an imputation process was carried out to reduce any bias that might be brought in due to missing values. The values were imputed using Hot Deck procedure. The steps involved categorizing the respondents in the sample into similar subgroups based on the variables sex, age group, ethnic group and household size. Missing data for respondents in any subgroup were then randomly replaced with comparable data from respondents in the same subgroup. The values were only imputed for those who reported having savings or are in debts. The imputation was done separately for 1995, 2000 and 2005 before aggregating data from the three waves into one data set.

The main independent variables used in the analysis are annual income, marital status, employment status, presence of children aged 16 or under, housing tenure, financial status and household size, as these variables are important in predicting both the existence and level of wealth (see for example Kan and Laurie (2010)). Also, other variables such as sex and year of data collection (wave) are included for control. Each variable in the data was observed at three time points (1995, 2000 and 2005). Using a long format type of data set in STATA 11, the analysis was done at the individual level. Before estimating the models, the data was introduced as a panel data set so that the multiple observations per person are linked to one case rather than being treated as different cases. However, clustering and stratification were not specified in the analysis as STATA –like many statistics software– does not allow this while estimating a panel data model. Therefore, this may lead to under

[6] Kan and Laurie (2010) reported that in 1995, 2000 and 2005 waves of the BHPS, the proportion of missing values for the amount held in savings were 18%, 21% and 21%respectively while the equivalent proportions for debts were 3%, 5% and 4%.

estimating the standard errors of estimates in this analysis and coefficients may seem more precise than what they actually are since clustering and stratification were not specified. However, although this may not lead to precise estimates, any differences between the produced estimates will be due to the difference between the two sets of weights since the modeling strategy is held constant. Therefore, it is possible to draw a conclusion on whether the two sets of weights are different and decide upon the optimum set. Furthermore, a conservative approach about accepting significance levels was implemented during results interpretation. This approach implies considering only the highly significant estimates and classifies estimates with low significance levels as not being significant.

Two random effects logistic regression models were used to estimate the determinants of having money in savings or being in debts respectively. However, each model was estimated twice using the two different longitudinal sets of weights. Similarly, to model the amounts of savings and debts, two random effects OLS regression models were estimated in which every model was estimated two times using the two sets of longitudinal weights.

With two different sets of longitudinal weights (one based on the respondents at waves 1, 5, 10 and 15 and the other based on the respondents at all waves up to wave 15), eight models were estimated as each set of weights is used to estimate the main four models.

The main idea is to assess the change on the regression coefficients when varying weight adjustments procedures. In particular, the point of interest is to identify the influence of creating non-response longitudinal weights based on the consideration of combination of waves with the same module of questions.

9. Results

9.1 Response propensity

Descriptive Findings

Table 1.1 shows the number of respondents, number of non-respondents, response rates and non-response rates in waves 5, 10 and 15 and in all waves up to wave 15 respectively. These rates were calculated based on respondents at wave 1. The response rate in the subset of waves (50.08%) is 4.7% greater than the response rate in all the 15 waves (45.41%). This difference in response rates is caused by 478 (4.7%) respondents who took part in waves 5, 10 and 15 but failed to respond in at least one other wave between 1 and 15. This result indicates that if a weighting adjustment that is based on respondents from all waves up to wave 15 is used for analysis of data from waves 1, 5, 10 and 15 it will assign a weight of zero to 478 respondents. Consequently, this approach, which corresponds to use of the BHPS wave 15 longitudinal weight, the only weight on the public use data file that could be used for this analysis, results in a loss of 4.7% of the sample that could be used for analysis from waves 1, 5, 10 and 15.

In contrast, using weighting adjustment that is based on response to waves 1, 5, 10 and 15, in the analysis of data from waves 1, 5, 10 and 15 will take into account these 478 respondents since the weighting model, in this case, identifies them as respondents.

Therefore, in our analysis, it seems reasonable to expect more precise estimates from using weights based on response to waves 1, 5, 10 and 15 than from using weights created from response to all of the 15 waves as the sample is larger (by 478 respondents) with the former set of weights.

As for the bias, even though the sample is larger, it cannot be established that using weights created from waves 1, 5, 10 and 15 will result in less biased estimates than weights created from all of the 15 waves. This is because the higher response rate in waves 1, 5, 10 and 15, which is associated with a larger sample drawn from this combination, does not necessarily reduce the potential bias in estimates. The bias may only be reduced if the additional 478 respondents are similar, in their main characteristics, to non-respondents.

Thus, in our comparison between estimates resulting from the two sets of weights, the analysis will focus on levels of precision rather than bias reduction.

Table 2 gives the proportions of sample members who maintained response and those who failed to maintain response in our two wave-combinations (1, 5, 10 and 15 and in all waves up to wave 15) by main characteristics from wave 1: sex, age, health status, work status, housing tenure and presence of children in household. For example, considering the first wave-combination, 47.25% of the men who responded at wave 1 also responded at waves 5, 10 and 15. Considering the second wave-combination, 42.44% of the men who responded at wave 1 also responded at all waves up to 15. In general, men are less likely to maintain survey response than women in both wave-combinations.

As for age and health status, both wave combinations reflect similar results. Those who are in the oldest age group (56 and over) are more likely to fail maintaining response compared to those who are in younger age groups (35 or below and between 35 and 56). Also, sample members who reported good health status seem to maintain response more than those with bad health status.

Similarly, sample members who are employed, own their accommodation and have children, registered lower rates of non-response in both wave-combinations than those who are unemployed, non-owners of their accommodation and do not have children respectively. These results reflect what might be expected in the association between the response propensity and the characteristics of individual/household from wave 1. Namely, the response propensity is expected to be positively correlated with being female, middle aged, having a good health, being employed, owning property, and the presence of children in the household.

Table 1.1 Number of respondents and non-respondents at waves 5, 10 and 15 and in all waves up to 15

	Respondents	**Non-respondents**	**Total**
Waves 5, 10 &15	5,132 (50.08%)	5,116 (49.92%)	10,248
All waves up to 15	4,654 (45.41%)	5,594 (54.59%)	10,248
Difference	478		

Table 2 Percentages of maintained and unmaintained response in the two wave-combinations by main characteristics of respondents at wave 1.

Variables from wave 1	Wave 1, 5, 10 & 15		All waves up to 15	
	Response	Non-response	Response	Non-response
Sex				
Men	47.25%	52.75%	42.44%	57.56%
Women	52.59%	47.41%	48.06%	51.94%
Age				
Aged 35 or below	54.44%	45.56%	48.43%	51.57%
Aged 36 to 55	58.47%	41.53%	53.21%	46.79%
Aged 56 or above	34.72%	65.28%	32.52%	67.48%
Health				
Good health	51.43%	48.57%	46.75%	53.25%
Bad health	35.35%	64.65%	30.81%	69.19%
Work status				
Employed	56.87%	43.13%	51.81%	48.19%
Unemployed	40.15%	59.85%	36.07%	63.93%
Housing tenure				
Owners	54.41%	45.59%	49.68%	50.32%
Non-owners	39.71%	60.29%	35.21%	64.79%
Presence of children				
Children in household	58.15%	41.85%	52.40%	47.60%
No children in household	46.00%	54.00%	41.88%	58.12%
Overall	50.08%	49.92%	45.41%	54.59%

Analytical Findings

The models in table 3 look into the relationship between the response propensity in waves 1, 5, 10 and 15 and in all waves up to wave 15, and characteristics related to respondents, household, interviewer and the interview observed at wave 1. The point of interest is to use the estimated probabilities in the two models to create the two sets of non-response weights. However, it is also useful to understand the difference in the tendency to responding in the two combinations of waves as based on variables from wave 1 since this difference leads to a dissimilarity in the two sets of weights we are interested in.

As shown in table 3, in the two wave-combinations, the response propensity is highly associated with gender, health status, work status and housing tenure. For example, in both combinations of waves, females are associated with the propensity to respond more than males ($\hat{b}_1$ = 0.550, $p < 0.01$; $\hat{b}_2$= 0.557, $p < 0.01$). Also, homeowners and those who report being employed at wave 1 are more likely to respond in both subsets of waves than non-owners and unemployed respondents respectively ($\hat{b}_1$ = 0.304, $p < 0.01$; $\hat{b}_2$ = 0.282, $p < 0.01$ and $\hat{b}_1$ = 0.567, $p < 0.01$; $\hat{b}_2$ = 0.576, $p < 0.01$). Additionally, both models show a highly significant negative association between bad health status and response ($\hat{b}_1$ = -0.288, $p < 0.01$; $\hat{b}_2$= -0.330, $p < 0.01$). These results are in line with the descriptive results in table 1.2 and the literature on attrition suggesting low response among men, respondents of bad health and home non-owners (Nicoletti and Buck, 2004; Lepkowski and Couper, 2002; Becketti *et al.*, 1988; Fitzgerald *et al.*, 1998; Watson, 2003; Nicoletti and Peracchi, 2005). As for work status, it is known to be problematic when trying to understand its relationship with the response propensity (Watson and Wooden, 2004). While some studies showed that the response propensity is high among unemployed respondents (e.g. Watson, 2003; Nicoletti and Peracchi, 2005), other studies support the finding in table 1.3 where there is a higher tendency to response among respondents who observed as employed in wave 1 (Gray *et al.*, 1996; Lepkowski and Couper, 2002). The explanation for this result might encompass the fact that employed respondents are more likely to be geographically immobile and hence easier to follow over time.

The response propensities in the two sets of waves are also significantly correlated with other individual/household and interviewer/interview characteristics from wave 1. For example,

Ethnicity: ethnicity is included in the model as a dichotomous variable specifying if respondent is white or non-white[7]. In both models, white respondents are more likely to respond compared to other ethnic groups ($\hat{b}_1$ = 0.580, $p < 0.01$; $\hat{b}_2$= 0.563, $p < 0.01$). This result is verified by the literature indicating that ethnic minority groups show lower tendency to response (Gray *et al.*, 1996; Lepkowski and Couper, 2002).

Education: in both models education is represented by a dummy variable indicating whether the respondent has a General Certificate of Education (GCE) level A-C or above. The results in table 3 show that those who have a GCE or any higher qualification, at the

[7] The ethnic minority groups in the British Household Panel Survey sample are too small to allow for valid analysis of different ethnic groups.

start of the survey, are more likely to carry on giving interviews in both subsets of waves ($\hat{b}_1$ = 0.276, p < 0.01; $\hat{b}_2$= 0.295, p < 0.01). This result is confirmed by a number of attrition studies (e.g. Watson, 2003; Gray *et al.*, 1996; Lepkowski and Couper, 2002).

Having a second job: although being employed at wave 1 is positively associated with the response propensity, the association is found to be negative if respondents have a second job ($\hat{b}_1$ = -0.426, p < 0.01; $\hat{b}_2$= -0.387, p < 0.01). The explanation for this may be that respondents with more than one job are harder to contact since they are less likely to be at home than respondents who have only one job or no job at all. Also, even if contact is made, it is more difficult to obtain cooperation from respondents who have more than one job as it is a trade-off of their leisure time.

Gender of interviewer: in the two models, respondents who were interviewed by women in the first wave of the study seem to maintain cooperation more than those who were interviewed by men ($\hat{b}_1$ = 0.180, p < 0.05; $\hat{b}_2$= 0.183, p < 0.05).

Turning to the differences between the two models, overall, the response propensity in waves 1, 5, 10 and 15 and in all waves up to wave 15 seems to associate similarly with most of the variables. However, some variables appear to be more important in one subset of waves than in the other. For instance,

The presence of children in household: this variable was included in the model under the assumption that the presence of children in a household increases the chance of contacting the household. However, this is only significant in the second model ($\hat{b}_1$ = 0.104, p > 0.10;$\hat{b}_2$ = 0.135, p < 0.05). This model shows that respondents from households with children are more likely to respond than respondents from households with no children.

Age: age is known to be a good predictor of the response outcome in most of the non-response literature (e.g. Uhrig, 2008). Nonetheless, in this analysis, it is only significant in the second model. The model shows that the increase in age is negatively associated with the response propensity ($\hat{b}_1$ = -0.002, p > 0.10; $\hat{b}_2$ = -0.006, p < 0.05).

Income: annual income could be viewed as an indicator of financial ability. The argument is that respondents with higher income are more likely to respond. The results from our models are in line with this argument. However, income is more significant in the second model ($\hat{b}_1$ = 0.001, p < 0.05; $\hat{b}_2$ = 0.001, p < 0.01).

Number of employed individuals in household: while the number of household members can associate positively with the probability of making contact with the household, the chance of non-contact is higher if more household members are in full-time employment. The results from our two models are in line with this assumption; however, this is more significant in the first model than the second one ($\hat{b}_1$ = -0.068, p < 0.01; $\hat{b}_2$ = -0.059, p < 0.05).

Savings: this was included as a binary variable (has savings=0 and has no savings=1). Although having money in savings is positively associated with responding in both models, the variable is more significant in the second model ($\hat{b}_1$ = -0.104, p < 0.05; $\hat{b}_2$ = -0.135, p < 0.01).

Presence of others during interview: if others are not present during the first interview, respondents tend to discontinue participating in the survey. This is likely to be because the presence of others is correlated with the possibility of making contact in subsequent waves (Uhrig, 2008). However, in this analysis, absence of others during the first interview is not significant in the first model ($\hat{b}_1$ = -0.069, p >0.10; $\hat{b}_2$= -0.100, p < 0.05).

Region: region is included in the models as a number of dummies with London as the omitted category. Both models give similar results about region in which the majority of the areas are more likely to respond than London. This result is confirmed by the findings of Uhrig (2008) in his analysis of BHPS attrition. Uhrig reported that the South-East, South-West, East Anglia and the North-East are more likely to respond than London. One difference between our two models, however, is that the South-East is not significant in the second model ($\hat{b}_1$ = 0.147, p < 0.10; $\hat{b}_2$= 0.128, p > 0.10).

Moreover, there are a few considerable differences in magnitude among coefficients in the two models. Thus, it is plausible to expect that a set of non-response weights generated from the first model would be different, in both weights sizes and weights variance, from a set of non-response weights generated from the second model.

Table 3 Logistic regression models of the response propensity in waves 1, 5, 10 and 15; and in all waves up to wave 15 based on information from wave 1.

	Model of response propensity for waves 1, 5, 10 and 15	Model of response propensity for all waves up to wave 15
Female	0.550***	0.557***
Employed	0.567***	0.576***
Employed female	-0.213**	-0.196**
White	0.580***	0.563***
Bad health	-0.288***	-0.330***
Household size	0.014	0.014
Household with children	0.104	0.135**
Home owner	0.304***	0.282***
Age	-0.002	-0.006**
Annual income/1000	0.001**	0.001***
Number of age 75 or over in household	-0.907***	-0.876***
Number in employment in household	-0.068***	-0.059**
Single person household	-0.224***	-0.222***
Has GCE qualification or above	0.276***	0.295***
Having a second job	-0.426***	-0.387***
Has no savings	-0.104**	-0.135***
Living in a flat	-0.245***	-0.279***
Based in business premises	0.159	0.009
Living in a bedsit	-0.406	-0.571
Living in other housing type	-0.154	-0.080
Interviewed by a female	0.180**	0.183**
Others not present when interviewed	-0.069	-0.100**
Lives in South-East	0.147*	0.128
Lives in South-West	0.192**	0.206**
Lives in East Anglia	0.656***	0.539***
Lives in the Midlands	0.029	0.062
Lives in the North	0.138*	0.142*
Lives in Wales	0.094	0.056
Lives in Scotland	-0.122	-0.157
Constant	-0.880***	-1.192***
N	10248	10248
Pseudo R^2	0.074	0.081

Note: The entries are marginal effects on the log-odds. The reference categories of the categorical independent variables in the two models are male, unemployed, non-white, good health, household with no children, not a house owner, multi-person household, does not have a GCE or higher degree, having no second job, has savings, living in a house, interviewed by a male, others present when interviewed and lives in London respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

9.2 Results from multivariate analysis of savings and debts

The models in tables 4 and 5 investigate the factors associated with the possession of savings and debt and the amount held in these in the years 1995, 2000 and 2005 respectively. Each model is estimated twice using our two sets of weights. The issue of interest here is to examine whether the two sets of weights lead to different results.

1. Possession of Savings and Debts

As seen in table 4, in all models, the possession of savings and/or debts is significantly associated with gender, age, financial situation, housing tenure, work status, and income. For example, women are more likely than men to have savings and debts ($\hat{b}_1 = 0.208$, $p < 0.01$; $\hat{b}_2 = 0.201$, $p < 0.01$; $\hat{b}_3 = 0.134$, $p < 0.05$; $\hat{b}_4 = 0.147$, $p < 0.05$), meanwhile, those who are out of the labour force are less likely to have savings and debts ($\hat{b}_1 = -1.129$, $p < 0.01$; $\hat{b}_2 = -1.121$, $p < 0.01$; $\hat{b}_3 = -0.963$, $p < 0.01$; $\hat{b}_4 = -0.992$, $p < 0.01$).

Focusing on the difference between the coefficients arrived at via the two sets of weights, the weights resulted in very similar fitted models. However, there are a few differences between the coefficients. For instance, in models concerned with savings, having a second job and being unemployed are significant in the first but not the second model ($\hat{b}_1 = -0.246$, $p < 0.01$, $\hat{b}_2 = -0.089$, $p > 0.05$; $\hat{b}_1 = -0.116$, $p < 0.05$, $\hat{b}_2 = -0.152$, $p > 0.05$). This is clearly showing the effect of the increase in the sample size used to estimate the first model on these particular variables (recall that the sample size used to estimate the first model is bigger by 478 respondents). The additional 478 respondents changed the sample in the first model both in terms of size and composition. This is especially interesting for the variable unemployed as its coefficient is nearer to zero, but still significant, in the first model than in the second model where it does not appear to be significant. In other words, using a weights adjustment method based on respondents in all waves which is associated with the loss of 478 respondents in the sample, results in underestimating the importance of having a second job and being unemployed.

As for debts, the difference appear with the variable 'having a dependent child' which is only significant when weights based on waves 1, 5, 10 and 15 are used to estimate the model ($\hat{b}_3 = 0.121$, $p < 0.05$; $\hat{b}_4 = 0.105$, $p > 0.05$).

Table 4 Random effects logistic regression models of possession of savings and debts.

	Having Savings		**Having Debts**	
	Using weights based on waves 1, 5, 10, and 15	**Using weights based on all waves up to wave 15**	**Using weights based on waves 1, 5, 10, and 15**	**Using weights based on all waves up to wave 15**
Year 2000	0.057	0.080	0.037	0.060
Year 2005	0.025	0.039	0.130**	0.116**
Female	0.208***	0.201***	0.134**	0.147**
Age	-0.007***	-0.008***	-0.040***	-0.041***
Financially okay	-0.841***	-0.832***	0.436***	0.468***
Having financial deficits	-2.384***	-2.439***	1.400***	1.401***
Mortgage payer	-0.117**	-0.122**	1.025***	0.995***
Council tenant	-0.416***	-0.422***	0.961***	0.948***
Private renter	-0.477***	-0.509***	0.922***	0.886***
Having a second job	-0.246***	-0.089	-0.466***	-0.242***
Having a dependent child	-0.302***	-0.297***	0.121**	0.105
Living with partner	0.001	-0.012	0.024	0.062
Member of a large household	-0.473**	-0.478**	-0.282	-0.247
Unemployed	-0.116**	-0.152	-0.234***	-0.270***
Out of the labour force	-1.129***	-1.121***	-0.963***	-0.992***
Annual income/1000	0.018***	0.019***	0.004**	0.004**
Constant	0.894***	0.933***	0.172	0.183
N	5132	4654	5132	4654

Note: The reference categories of the dependent variables are having no savings and having no debts respectively. The reference categories of the categorical independent variables in the models are year 1995, male, having a good financial situation, outright owner, has no second job, has no dependent child, not living with a partner, having a small household and employed respectively. ** $p < 0.05$, *** $p < 0.01$.

2. Amount of Savings and Debts

Table 5 shows that the amounts of savings and debts significantly depend on gender, age, financial situation, housing tenure, work status and income. For instance, mortgage payers have lower level of savings and debts than outright house owners ($\hat{b}_1$ = -0.096, $p < 0.01$; $\hat{b}_2$= -0.105, $p < 0.01$; $\hat{b}_3$ = -0.124, $p < 0.01$; $\hat{b}_4$= -0.109, $p < 0.01$). Also, income is positively correlated with the amounts held in savings and debts ($\hat{b}_1$ = 0.009, $p < 0.01$; $\hat{b}_2$= 0.009, $p < 0.01$; $\hat{b}_3$ = 0.009, $p < 0.01$; $\hat{b}_4$= 0.008, $p < 0.01$).

Turning to the differences amongst coefficients, regarding the amount of savings, having a second job is significant in the first model but not in the second model ($\hat{b}_1$ = 0.038, $p <$ 0.05; $\hat{b}_2$= 0.011, $p > 0.05$). Yet again, the explanation for this is that the first model uses extra respondents and hence gains more precision for its estimates.

Considering the level of debts, being financially okay, having financial deficits, and being unemployed are seem to be more significant when weights based on waves 1, 5, 10 and 15 are used to estimate the model ($\hat{b}_3$ = -0.040, $p < 0.01$, $\hat{b}_4$= -0.043, $p < 0.05$; $\hat{b}_3$ = -0.074, $p < 0.01$, $\hat{b}_4$= -0.102, $p < 0.05$; $\hat{b}_3$ = -0.092, $p < 0.01$, $\hat{b}_4$= -0.074, $p < 0.05$). Additionally, having a second job does not appear to be significant if estimated using weights based on all waves up to wave 15 ($\hat{b}_3$ = 0.071, $p < 0.05$; $\hat{b}_4$= 0.024, $p > 0.05$).

Table 5 Random effects OLS regression models of the amount of savings and debts.

	Savings		**Debts**	
	Using weights based on waves 1, 5, 10, and 15	**Using weights based on all waves up to wave 15**	**Using weights based on waves 1, 5, 10, and 15**	**Using weights based on all waves up to wave 15**
Year 2000	0.008	0.007	-0.002	-0.008
Year 2005	-0.007	-0.013	-0.026	-0.020
Female	-0.045***	-0.040***	-0.226***	-0.224***
Age	-0.001**	-0.001**	0.006***	0.006***
Financially okay	-0.107***	-0.104***	-0.040***	-0.043**
Having financial deficits	0.035	0.039	-0.074***	-0.102**
Mortgage payer	-0.096***	-0.105***	-0.124***	-0.109***
Council tenant	-0.094***	-0.100***	-0.380***	-0.383***
Private renter	-0.048**	-0.062**	-0.183***	-0.185***
Having a second job	0.038**	0.011	0.071**	0.024
Having a dependent child	-0.018	-0.017	-0.014	-0.007
Living with partner	0.046***	0.047***	-0.010	-0.027
Having a large household	0.046	0.047	0.070	0.044
Unemployed	-0.006	0.005	-0.092***	-0.074**
Out of the labour force	0.058***	0.057***	0.027	0.032
Annual income/1000	0.009***	0.009***	0.009***	0.008***
Constant	-1.980***	-1.985***	0.836***	0.842***
N	5132	4654	5132	4654

Note: The dependent variables in all the models are transformed to the natural logarithm. The reference categories of the categorical independent variables in the models are year 1995, male, having a good financial situation, outright owner has no second job, has no dependent child, not living with a partner, having a small household and employed respectively. ** $p < 0.05$, *** $p < 0.01$.

In sum, the two models lead to similar results in showing no big differences between the coefficients in the two models. Thus it can be concluded that weight adjustments based on the respondents in waves 1, 5, 10 and 15, lead to similar results as weight adjustments based on the respondents at all waves up to wave 15, when analysing wealth data from waves 1, 5, 10 and 15 from BHPS. However, the latter set of weights produces less precise results for some of the estimates. This difference in precision results because, when using weights based on respondents from all of the 15 waves with data from waves 1, 5, 10 and 15, some respondents will be lost as the weighting model, in this case, identifies response as responding in all of the 15 waves.

10. Conclusion

A longitudinal type of survey is the best design for a thorough understanding of the change in dynamic populations. However, like all surveys, a longitudinal survey must undergo a level of non-response that may negatively affect its excellent reward. In recent decades, rates of non-response have risen in most survey research. Perhaps this is why survey researchers have become more wary about non-response and more convinced regarding the use of statistical weights as adjustment for its negative consequences. Weighting is an effective way to deal with non-response at the analysis stage if it is used appropriately. However, the customary use of non-response weighting that is currently implemented in longitudinal surveys may still be insufficient. But it could be well promoted if a few considerations were taken into account.

In longitudinal surveys non-response is not a one-off event (Watson and Wooden, 2009) it is rather dynamic and can take different patterns among different sub-periods of time during the life of the panel. Therefore, different combinations of data collection points might suffer from different sizes of non-response error. Consequently, the amount of error can vary not just between estimates of different parameters but also within estimates of the equivalent parameter constructed at different points in time in the same survey. This variation might be due to changes in the responding status among different combinations of waves (resulting in a change in the sample size and/or the sample composition). Thus, an ordinary weighting strategy, which does not take into account the changes in the responding sample between wave-combinations, can only deal with the fixed part of non-response error. Instead, a weighting strategy that takes into account the change in the responding sample between wave-combinations and involves estimating more weighting models to produce more than one set of weights can tackle the fixed as well as the variable part of non-response error.

This research demonstrates the limitation of a single weighting strategy in dealing with non-response error in all subsets of waves. In a single weighting strategy, the weighting model is estimated by reference to the responding status in all of the waves. Thus, it may lead to an unnecessary loss of respondents in any analysis that uses a sample from just a subset of waves. This loss of respondents, depending on its level, can result in less precise estimates which may increase the total survey error. The substantive comparison between the models in this paper shows that using non-response weights - equivalent to the ones provided to BHPS users- from such weighting strategy to analyse wealth data from waves 5, 10 and 15 from the BHPS does not take into account 478 respondents who are actually present in this combination of waves. Moreover, an alternative set of weights that is designed specifically to consider these 478 additional respondents, results in different outcomes (more precision on some estimates) compared to the set of weights from the single weighting strategy.

Although the alternative strategy increases the precision in estimates (in comparison with the standard strategy with a single weighting model) by using 478 additional respondents, the analysis here does not ascertain whether or not it also reduces bias. For bias to be reduced, the additional 478 respondents, who skipped responding at some waves (i.e. wave non-respondents), would have to be more similar, in their characteristics, to non-respondents to waves 1, 5, 10 and 15 than to respondents. However, we currently have no evidence to support this claim.

A single weighting strategy do take care of non-response error on several estimates, but clearly fails in tackling the error introduced in other estimates due to the loss of information. In turn, creating a set of weights for respondents in every possible combination of waves is not practical and may be unachievable sometimes, in particular if a large number of waves are conducted. But a limited number of subsets of weights could be produced for significant wave-combinations. A challenge then is to identify specific wave-combinations that are worth creating a subset of weights.

In this paper, the identification of the wave-combination of interest is driven from the analytical interest perspective. The selected wave-combination collects information about the same subject. Such wave-combinations are common in longitudinal surveys where some surveys include certain questions in specific waves. Thus, analyses conducted on data drawn from these waves may benefit from a weighting that is based on these waves. The analysis in this paper shows that weighting based on this approach results in more precise estimates. Although, in broad terms, this approach of weighting gave similar results to the single weighting approach, having different results on some estimates indicates how a single weighting model might be sub-optimal. In the present research, the introduced weighting resulted in avoiding the loss of 478 respondents. In other situations, the suggested weighting approach may help avoiding the loss of a bigger proportion of respondents. In which case, the new weighting will not just substantially improve estimates but could also produce results with different significance levels compared to the results from the single weighting approach. This is particularly more likely in surveys that are designed to exist for a long time (i.e. will have more waves) such as Understanding Society in the UK. Thus, in these surveys, this approach of weighting is worth considering.

Additionally, other features of longitudinal surveys may push for different types of considerations to be taken into account too. For example, to enhance the accuracy of survey estimates, survey organisations sometimes add extra information to the original sample. For instance, two samples (from Scotland & Wales) were added to BHPS in wave 9. Also, an additional sample (from Northern Ireland) is added at wave 11. Thus, for BHPS, providing subsets of weights for waves 9 onwards, and 11 onwards might be of interest. Moreover, Lynn and Kaminska (2010) put forward other creditable considerations when deciding on wave-combinations for weighting. However, although each of the suggested considerations maybe beneficial, at present, there is no approach that guides all of –or at least some of- these considerations simultaneously for rational selection of wave-combinations for weighting. Instead, each consideration, if used, will be used solely. Meanwhile, weighting can benefit greatly by combining all or some of these considerations. Thus, a challenge for a future research is to unite all these considerations into one constructive method that would guide the process of selecting wave-combinations.

In any case, the choice of specific wave-combinations for weighting should be guided by a rule that takes into account two issues:

(a) The subsample drawn for analysis from any chosen combination of waves should be considerably different (in size and composition) from the subsamples in other wave-combinations.
(b) The selected combination of waves should be usable for analysis that achieves the objectives of the survey.

References

Alreck, P., and Settle, R. (1995) "The Survey Research Handbook", New York: McGraw-Hill.

Bates, N., Dahlhamer, J., and Singer, E. (2008) "Privacy Concerns, Too Busy, or Just Not Interested: Using Doorstep Concerns to Predict Survey Nonresponse", *Journal of Official Statistics*, vol 24, pp. 591-612.

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. (2009) "Applied Survey Methods: A Statistical Perspective*",* New Jersey: 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.

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

Cochran, W. (1977) "Sampling Techniques"*,* John Wiley and Sons.

Curtin, R., Stanley, P., and Singer. E. (2000) "The Effects of Response Rates Changes on the Index of Consumer Sentiment", *Public Opinion Quarterly,* vol. 64, pp413-28.

Deville, J. C., and Sarandal. C. E. (1992) "Calibration Estimators in Survey Sampling*",* *Journal of the American Statistics Association*, vol. 87, pp. 376-382.

Dillman, D. (2007) "Mail and internet surveys: The tailored design method*",* New York: John Wiley & Sons Inc.

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.

Gouskova, E. (2001) "PSID Technical Report: The 2002 PSID Child Development Supplement (CDS-II) Weights*"*. [online] available from http://psidonline.isr.umich.edu/data/weights/ [Accessed 19th December 2010].

Grau, E., Potter, F., Williams, S., and Diaz-Tena, N. (2006) "Nonresponse Adjustment Using Logistic Regression: to weight or not to weight"*,* New Jersey: Mathematica Policy Research, Inc. [online] available from http://www.amstat.org/sections/srms/proceedings/y2006/Files/JSM2006-000717.pdf [Accessed 19th October 2010].

Gray, R., Campanelli, P., Deepchand, K., and Prescott-Clarke, P. (1996) "Exploring Survey Non-response: the effect of attrition on follow-up of the 1984-85 Health and Life Style Survey", *The Statistician*, vol. 45, pp. 163-183.

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

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

Groves, R. (2006) "Nonresponse Rates and Nonresponse Bias in Household Surveys", *Public Opinion Quarterly*, vol. 70, pp. 646-675.

Groves, R., and Peycheva, E. (2008) "The Impact of Nonresponse Rates on Nonresponse Bias: A meta Analysis", *Public Opinion Quarterly*, vol. 72 (2), pp. 89-167.

GTSS (Global Tobacco Surveillance System). (2009) "Global Adult Tobacco Survey (GATS): sample weights manual". [online] available from www.cdc.gov/tobacco/global/gats. [Accessed 5th October 2010].

Kan, M., and Laurie, H. (2010) "Savings, Investments, Debts and Psychological Well-being in Married and Cohabiting Couples", *ISER Working Papers Series,* No. 2010-42. Colchester: University of Essex.

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

Kreuter, F., and Krieger, U. (2009) "Analyzing Contact Sequences in Call Record Data. Potential and Limitations of Sequence Indicators for Nonresponse Adjustments in the European Social Survey", *Journal of Official Statistics*, vol. 25:2, pp. 203-226.

Kroh, M. (2009) "Data Documentation: sample sizes and panel attrition in the German Socio Economic Panel (1984 until 2008)", Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).

Lepkowski, J., and Couper, M. (2002) "Nonresponse 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.139-157.

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

Lynn, P. (2003) "PEDAKSI: Methodology for Collecting Data About Survey Nonrespondents", *Quality and Quantity*, vol. 37, pp.239-261.

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 1
to 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 Nonresponse", (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., and Kaminska, O. (2010) "Criteria for Developing Non-response Weight Adjustments for Secondary Users of Complex Longitudinal Surveys", *2010 International Workshop on Household Survey Non-Response.* Nürnberg.

Lynn, P., Kaminska, O., and Goldstein, H. (2014, forthcoming) "Panel Attrition: How Important is it to Keep the Same Interviewer?", *Journal of official statistics*, (forthcoming).

Madow, W., Nisselson, H., and Olkin, I. (1983) "Incomplete Data in Sample Surveys", New York: Academic Press Inc.

Massey, J. T. (1988) "An overview of telephone coverage", in Groves, R., Biemer, P., Lyberg, L. E., Massey, J. T., Nicholls, W. L., and Waksberg. J, *Telephone Survey Methodology.* New York: John Wiley & Sons, Inc.

Mohadjer, L., and Choudhry, G. (2001) "Adjusting for Missing Data in Low-Income Surveys", in Ploeg, M., Moffitt, R., and Citro, C. *Studies of Welfare Populations: Data Collection and Research Issues.* Washington: The National Academies Press.

Merkle, D., and Murray, E. (2002) "Nonresponse in Exit Polls: A comprehensive Analysis", in Groves, R., Dillman, D. A., Eltinge, J., and Little, R. J. A., *Survey Nonresponse*. New York: Wiley.

Nicoletti, C., and Buck, N. (2004) "Explaining Interviewee Contact and Co-operation in the British and German Household Panels", (pp. 143-166) in Ehling, M., and Rendtel, U. *Harmonisation of Panel Surveys and Data Quality*. Wiesbaden: Statistisches Bundesamt.

Nicoletti, C., and Peracchi, F. (2005) "Survey Response and Survey Characteristics: microlevel evidence from the European Community Household Panel", *Journal of the Royal Statistical Society*, Series A, vol. 168, pp.763-781.

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.

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

Scheaffer, R., Mendenhall, W., and Ott, R. (1996) "Elementary Survey Sampling", California: Wadsworth.

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

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.

Taylor, M. (ed.). (2010) "British Household Panel Survey User Manual Volume A". Colchester: Institute for Social and Economic Research, University of Essex.

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.

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. (2004) "The HILDA Survey Four Years on", *Australian Economic Review*, vol. 37, pp.343-349.