Using Proxy Measures of Survey Outcomes in Post-Survey Adjustments: Examples from the European Social Survey (ESS)

Frauke Kreuter, Michael Lemay, Carolina Casas-Cordero
Joint Program in Survey Methodology, University of Maryland

Abstract

Sample survey nonresponse may introduce bias if nonrespondents are not missing at random. Post-survey weighting adjustments are one strategy for addressing survey nonresponse. To be effective, the variables used in the adjustment need to be highly correlated with both response propensity and with the survey variable of interest. Interviewer observations made during the field period are increasingly available for respondents and nonrespondents. Using data from the ESS this paper will examine the extent to which these observational data are correlated with response propensity and a selected set of survey variables.[1]

Keywords: Nonresponse Adjustment, Interviewer Observations, European Social Survey

1 Introduction

Sample survey nonresponse may introduce bias if nonrespondents are not missing at random. Strategies for increasing survey response rates include the use of respondent incentives, additional call attempts and refusal conversion protocols. Such efforts have had mixed success. Post-survey adjustments are another strategy for addressing survey nonresponse. To be effective in reducing bias and variance, the variables used in the adjustment need to be highly correlated with both response propensity and with the survey variable of interest (for a recent simulation see Little and Vartivarian (2005)). Easily available variables such as gender, sex, and race are traditionally used in nonresponse adjustment procedures and other post-survey adjustments. Unfortunately, these variables are not always closely related to response probability[2], and may be only loosely correlated with the outcome variables of interest. However, there are other variables available but rarely considered for weighting purposes, namely those collected for survey management and those that could be collected by interviewers with little extra expense. The present paper will examine the extent to which these other variables can be seen as successful candidates for nonresponse adjustments in the European Social Survey.

Two examples of observations for both respondents and nonrespondents - number of contact attempts and observed signs of incivility - may illustrate how these variables can be effectively used for nonresponse adjustments. In many surveys, variables such as the date and time of contact are part of the process data collected for survey management purposes. These variables are useful, as they may be related to participation propensities and outcome variables. For example, conditional on the use of a balanced call schedule, the information available from contact attempts can be used to create estimates for "time spent at home". "Time spent at home" is a predictor of the probability of making a successful final contact given a fixed field period. "Time spent at home" can likewise be highly correlated with the substantive outcome variable such as provision of after-school care for children. If people who are rarely at home and are therefore hard to contact are disproportionally missing from the survey sample, estimates the amount of parental after-school care will be biased. A nonresponse adjustment procedure that takes the number of contact attempts into account can therefore reduce bias for this particular outcome variable.

A second example involves data collected through interviewer observation. An in-person interviewer can, for example, record the physical and social characteristics of selected housing units, and of the neighborhoods where these are located. These characteristics may include the presence of graffiti, litter etc. - often referred to as signs of incivility, and are thought to be related to fear of crime (Sampson and Raudenbush, 1999). Fear of crime may, in turn, be one reason to refuse participation in an in-person survey (Groves and Couper, 1998). Thus, observed signs of incivility can predict both cooperation and survey measures in a survey on neighborhood involvement or fear of crime.

2 Background and Significance

Few, if any, surveys obtain responses from every element originally selected into the sample. Observed nonresponse rates vary widely across survey characteristics such as mode of data collection, survey organization or geographical area surveyed. Despite these variations, one general trend is noticeable - that of declining response rates (Schnell, 1997; Groves and Couper, 1998; De Leeuw and DeHeer, 2002; Curtin et al., 2005; Groves

[1]This paper is part of a larger project that resulted from a research seminar run by Robert Groves and Trivellore Raghunathan at the Joint Program of Survey Methodology (JSPM) and the Michigan Program of Survey Methodology (MPSM). We thank all seminar participants as well as Rainer Schnell and Richard Valliant for helpful comments.

[2]For a critical discussion of the notion of response propensity or response probability see Dalenius (1983).

and Heeringa, 2006).

Unfortunately, response rates have declined despite the increasing efforts of survey organizations and researchers to reach households and gain cooperation (Dillman et al., 2002). First, surveyors have made efforts to improve contact rates (for example, through a greater number of contact attempts, a longer data collection period or the use of multiple modes to establish contact - all of which are known to have a positive effect on response). Secondly, surveyors have also tried to increase cooperation through the use of, for example, incentives, advance notice of the survey request, follow-up procedures, and alternative response modes. However, neither set of techniques has stopped or reversed the downward trend in response rates. Assuming that these measures are at least somewhat helpful indicates that response rates would have decreased even further without these efforts. This means that survey designs that aim for a high response rates are subject to the increasing costs incurred by repeated efforts to obtain access to sample units and more involved attempts to address the concerns and hesitations of the sample members (Rogers et al., 2004; Groves and Peytcheva, 2006).

With the goal of removing bias due to nonresponse, survey researchers and survey organizations often create post-survey nonresponse adjustments; either model-based methods, or weighting methods. This paper focuses on weighting methods in which data for respondents whose characteristics are similar to the nonrespondents (based on covariates available for both) will have a higher weight to compensate for the units lost through nonresponse. Among those, weighting class adjustment is one approach, where all sample units–responders and nonresponders–are classified into cells based on characteristics that are predictors of whether a unit responds or not. Another method is the use of propensity scores. Such scores combine several characteristics for the unit of observation (e.g., person) into a single composite measure through a mathematical model that predicts someone's propensity to respond. In weighting class adjustments only a small set of categorical variables is typically used to ensure cell sizes are sufficient for a stable adjustment. A larger list of these variables as well as continuous measures can be incorporated into a model to predict propensity scores (Little and Rubin, 2002).

While nonresponse adjustment weights solve the nonresponse problem in theory, in practice their use can be problematic. First, for the adjustment models to be effective in reducing nonresponse bias they have to use variables that are related to both the probability of response as well as to key survey outcomes (Little, 1986; Kalton and Maligalig, 1991; Kalton and Flores-Cervantes, 2003; Little and Vartivarian, 2005; Groves, 2006). This is a challenging problem. While some correlates of the likelihood of responding are known, the mechanisms are not. Hence, it is difficult to predict which of the sample unit characteristics that are correlated with nonresponse are also (highly) correlated with any given key survey outcome. We therefore do not know for a given survey variable of interest if it will be affected by nonresponse bias or not.

Second, for 'sample based' nonresponse adjustments, the variables used in the adjustment need to be available for both respondents and nonrespondents.[3]. The sampling frame (the list from which the sample is selected) can sometimes provide useful information for the assessment of nonresponse bias, especially if the sampling frame is a population register or administrative list of some sort. In both cases the approach can only aim to correct for that part of the nonresponse bias that is explained by the variables used to build the weighting classes. The procedure requires the assumption of a strong correlation between the classes and both the survey measures and response propensity. In practice, such correlations are often rather modest. Many survey researchers are consequently wary of the ability of standard weighting techniques to adequately correct for unit nonresponse bias (Schnell, 1993; Lynn, 2003).

2.1 Importance of Paradata for Nonresponse

Largely untapped sources of information that could be used for adjustment are paradata, such as the interviewer observations and contact-record data collected for in-person surveys during the main field period. Adding these variables into the adjustment procedures is possible with little expense to the researchers and might help reduce nonresponse bias.

Recently, the U.S. Census Bureau implemented an automated system for collecting contact histories for computer-assisted personal interviews (Bates et al., 2006). Other government statistical agencies have started using similar procedures. In all of these cases, information is gathered primarily to be used by interviewers and field operations to tailor subsequent contact attempts with the potential respondent (Groves and Couper, 1998; Groves and McGonagle, 2001).

Several in-person surveys collect even more information through interviewer observation, for example the National Survey of Family Growth (NSFG), the British Crime Survey (BCS) or the European Social Survey (ESS). There information is collected on various neighborhood and housing unit characteristics, such as impediments to access to the unit (e.g. locked gates), number of housing units in the structure, evidence of children (e.g. toys visible) and evidence of the physical state of dwellings in the area. A similar set of measures was recorded by the interviewers and enumerators[4] for the German DEFECT study on fear of crime (Schnell and Kreuter, 2000a).

There are a couple of examples in which these more extensive data had been used to estimate survey participation. Lynn (2003) demonstrated how observations

[3]For 'Population based' adjustments only population totals for the adjustment variables are needed.

[4]Persons creating a list of addresses for the sampling frame.

on multiunit structures and door intercoms predict the amount of effort that is required to contact sample households in the British Crime Survey. Bates et al. (2006) used the contact information from the 2005 National Health Interview Survey (NHIS) to predict survey participation. Copas and Farewell (1998) successfully used the interviewer-assessed enthusiasm of sample members about participating in the British National Survey of Sexual Attitudes and Lifestyle as a predictor for the likelihood of response. The correlation with the likelihood of responding demonstrated in the above studies is only one aspect of good adjustment variables culled from contact histories and interviewer observations.

Fewer studies have examined the relationship between contact history data and the key survey variables of interest. To give some recent examples, Schnell and Kreuter (2000b) reported a correlation between the number of contact attempts and self-reported criminal victimization in several studies. Riede and Asef (2006) demonstrated the potential for nonresponse bias in the estimation of labor market participation as a function of contact attempts. More knowledge is available for the relationship between contact history and demographic characteristics of respondents, with 'late' responders being more likely to be younger, to live in smaller households, and to live in rental units or houses with access impediments (Groves and Couper, 1998; Duhart et al., 2001). The most extensive use of interviewer observations can be found in Raessler and Schnell (2004). In their study several adjustment methods were compared for a survey on fear of crime.

While the literature includes suggestions to expand the use of paradata in nonresponse adjustment (for example Raessler and Schnell (2004), Lynn (2003), and Kalton and Maligalig (1991))[5], survey researchers have so far not taken this step on a large scale. Despite successful demonstrations – for example Czajka et al. (1987, 1992), who investigated the use of propensity scores to account for late filers in the Survey of Income (Sol) – only a small number of household surveys have followed these examples.

2.2 Missing Pieces

Several important aspects of the use of interviewer observations and contact history data have yet to be clarified. First, the relatedness of these auxiliary variables to the entire set of variables in a given survey has often not been examined. Theory suggests these indicators work well for adjustments to some variables but not for others. The relative advantage needs to be examined. Second, given time and attentional limits, priorities must be set as to which observations an interviewer should make. While these data are relatively inexpensive to collect, their collection can interfere with the interview process, for example if the interviewer has to record observations between establishing cooperation and starting the actual interview. For some survey organizations even the collection of contact protocol data constitutes a noticeable burden. To gain cooperation with the collection of paradata and to reduce the interviewer burden, survey researchers need to know which call history data or interviewer observation is most useful for post-survey adjustment.

As part of a larger project this paper will address the first question using data from the European Social Survey (ESS). For the ESS we will examine the correlation between interviewer observations and a set of outcome variables, as well as the correlation between the interviewer observations and the response behavior. Estimates will be compared just using the selection weights and using nonresponse weights build from interviewer observations in addition to the selection weights.

3 European Social Survey

The European Social Survey (ESS) is a multi-purpose survey that takes place every two years. It is designed to measure and monitor changing social attitudes and values across Europe, identifying and interpreting trends over time. The first round of data collection took place in 2002 and its (mostly attitudinal) items can be grouped into the following topics: Media, Social Trust, Politics, Subjective Wellbeing, Immigration and Asylum as well as Citizen Involvement. A series of demographic questions concludes the questionnaire. The survey was conducted face-to-face and the average interview time of the core instrument was about one hour.

Great effort was put into standardization of field procedures across countries. To minimize fieldwork variation between the countries, all interviewers were instructed to make at least four personal visits to each sampling unit before a case was abandoned as non-productive. The interviewers were instructed to spread out those visits over at least two different weeks, and to include at least one attempt in the evening, as well as one during the weekend. The first contact with potential respondents, following a possible advance letter, had to be face-to-face. Once in contact with a household, however, interviewers were permitted to make appointments by phone (Philippens and Billiet, 2007)[6]. The participating countries were free to increase efforts beyond the prescribed rules[7].

Figure 1 summarizes the response rates. About half the countries participating in round one of the ESS came close to the specified target response rate of 70% (the median response rate is 65%). In all countries refusal-rates are higher than non-contact rates. The aim to keep

[5]For early discussions on the estimation of contactability see Politz and Simmons (1949).

[6]The restriction on making initial contact by telephone was relaxed for some countries. Analysis of the call records, points out that Switzerland, Sweden, Finland and Norway predominantly used telephone attempts to recruit respondents.

[7]Irish, Slovenian and Greek interviewers for example had to make at least five contact attempts, while Polish and Slovenian interviewers had to make at least two evening calls.

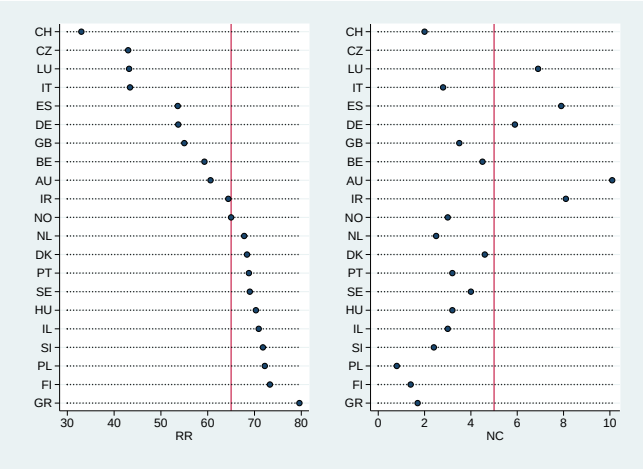

Figure 1: Response rate and non-contact by country (Source: Philippens and Billiet (2007), Table 1)

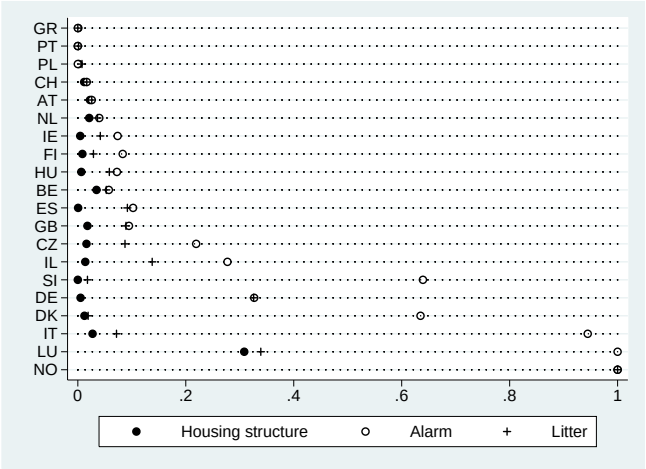

Figure 2: Proportion missing values for each item by country

non-contact rates to a minimum was achieved in most countries with the exception of Germany, Luxembourg, Spain, Ireland and Austria, where the non-contact rate is above 5%. The analysis by Philippens and Billiet (2007) shows a lack of contact attempts for most of those countries. The variation in response rates across countries has lead to increase interest among ESS researchers to possible nonresponse adjustments. Here we will examine the potential of interviewer observations as nonresponse adjustment variables.

3.1 Interviewer Observations

Interviewers of all countries provided information on a contact protocol. The contact protocol used for the European Social Survey data collection included for each contact attempt the ID of the interviewer handling the case, date and time of contact, mode of contact, result from the visit, in case of refusal the reasons for refusal, interviewer estimation of the likelihood of cooperation as well as age and gender of the respondent estimated by the interviewer. In addition the interviewer completed a housing and neighborhood characteristics form for each address.

The ESS interviewer observations of the housing unit and neighborhood can be grouped into three sets.

- The interviewers were first asked to record the type of housing structure the respondents live in. Interviewers distinguished between farm houses, various forms of single-unit and multi-unit housing, and other categories such as living in a trailer or a boat. For our purpose here, we created a dummy variable separating any form of multi-unit housing (including student apartments and sheltered housing) from the rest.

- The second set of observations made by the interviewer are observations on the housing unit itself. Based on outside visibility at the sampled address the interviewer is asked to mark the presence or absence of an alarm system, intercom or entry phone, security lights, porch (both open and closed), Beware of dog sign, bars or grills on any window.

- The third set of observations are about the area the sample address is in, such as: "In what physical state are the buildings or dwellings in this area?"; "In the immediate area, how common is litter or rubbish lying around?" and "How common is vandalism, graffiti or deliberate damage to property?" The interviewer was also asked to rate the physical state of the sampled address in comparison to other addresses in the immediate area.

Data quality is an issue of concern before we start our analysis. To use interviewer observations as a tool for assessing or adjusting for nonresponse bias, those observations should be made for all eligible sample units (respondents and nonrespondents).[8] If interviewer observations are missing, the mechanism that led to their missingness should be unrelated to response propensity.

Figure 2 shows the fraction of missing information on all eligible sample cases within each of the countries for one indicator out of each set of interviewer observations. In Figure 2 the countries are ordered by the average number of missingness across those three interviewer observations displayed here: a binary indicator whether it is a multi-unit structure or not, the presence or absence of an alarm system, and the presence or absence of litter in the immediate area.

As can be seen in Figure 2 data from Norway are all missing for those three variables in the publicly available datafile. The contact protocol data from Luxembourg

[8] A sample case was considered ineligible if the interviewer recorded 'ineligible' at one of the contact attempts, even if further contact attempts were made thereafter.

do not contain information on the presence or absence of an alarm system. They do however show high rates of missing data on the variables 'housing type' and 'litter'. In several countries such as Finland, Italy and Denmark, the completeness varies between measures, with the neighborhood observation being the most complete. For example, Denmark has around 1% missing data on the neighborhood measure and the information on the housing structure but around 60% missing for the housing unit observation (alarm). Missingness is correlated within the three groups of interviewer observations discussed earlier. Interviewer observations that describe the housing unit have in general higher missing data rates in the ESS contact protocol data.[9]

Interviewer observations in the ESS are complete for a small number of countries only. The observed correlations could be biased if the observed patterns of interviewer-observation missing data are not missing at random or missing completely at random. Greece, Poland and Portugal have the most complete set of interviewer observation. There are virtually no missing values on any of the interviewer observation for these countries. We will use these three countries for our example analysis here.

3.2 Study variables of interest

For our purpose we focus on a subset of survey questions that were identified to be (likely) correlates to the interviewer observations or response propensity, i.e. items about social involvement and more general social trust (Schnell, 1997; Couper et al., 1998; Groves and Couper, 1998; Abraham et al., 2006). For example "Do you think that you could take an active role in a group involved with political issues?" and "How interested would you say you are in politics? ", as well as two items on trust: "Please tell me how much you personally trust .... [the legal system], [the United Nations]". Items that have potential to be related to the interviewer observations are questions about fear of crime and victimization, for example "How safe do you or would you - feel walking alone in this area after dark?", "Have you or a member of your household been the victim of a burglary or assault in the last 5 years?". In addition we added one item from the health section "How is your health in general?" and one item related to in-house activities, the number of hours TV is watched.

4 Analysis

Our goal here is to determine the effect on ESS estimates when one or several of the interviewer observations are used for nonresponse (weighting class) adjustment. As a first step in our analysis we examined the correlations between each of the interviewer observations and the likelihood of being a respondent. All eligible sample cases for which interviewer observations are available are used for this analysis. The second step involves examining the correlations between interviewer observations and the key survey variables identified above. This correlation can only be estimated for respondents. In this paper we have to operate under the assumption that the correlation between the interviewer observation and survey variables is similar among the nonrespondents. Finally we compute the difference in point estimates for data weighted only with selection weights and those weighted with selection weights as well as a nonresponse adjustment weight created out of the interviewer observation.

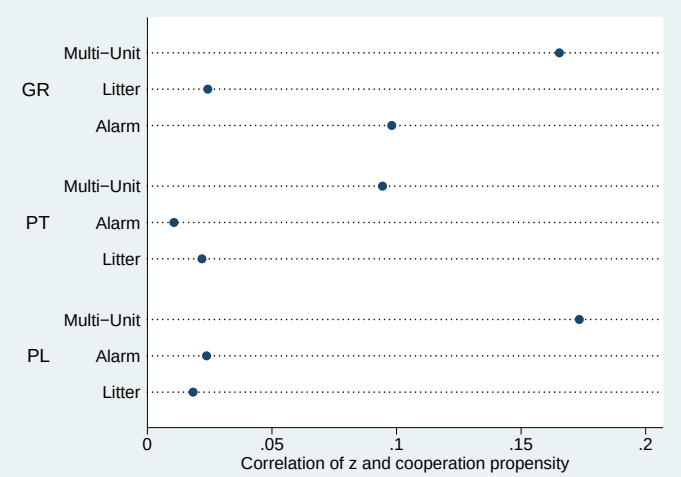

Figure 3: Correlation of three different interviewer observations with response for Greece, Portugal and Poland

4.1 Correlations

As mentioned before, in order for the data to be useful for nonresponse adjustment, interviewer observations must be related to both the propensity to be a respondent (p) and the survey variables of interest (Y).

Figure 3 gives three examples for the correlation of interviewer observations and the participation in the survey. Two aspects are noteworthy here. First, the (absolute) correlations are low. Second, the strength of the correlation varies not only by observation but also by country. The presence or absence of an alarm system for example shows a higher correlation (around 0.1) in Greece, but is below 0.05 in Portugal and Poland.

Figure 4 shows for the three selected countries, separately and jointly, box-plots for the correlation of the interviewer observations on the presence of litter, multi-unit structure, and alarm system with eight key outcome variables. Here we see again mostly very small correlations and variation across countries. What is important to note in addition is that the correlations also vary across survey variables.

[9] We hypothesize that for the interviewer observation of the housing unit the visual layout of the interviewer questionnaire could have led to a missing when the item was not present. Interviewer were asked to record for each item its presence of absence. In retrospect we can not distinguish between 'a true' missing value and a missing mark on 'no, absent'.

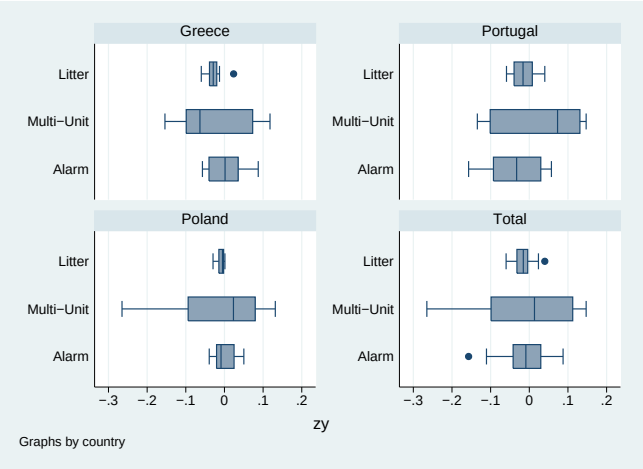

Figure 4: Box plots for the correlation between Z variables and a set of eight key survey variables

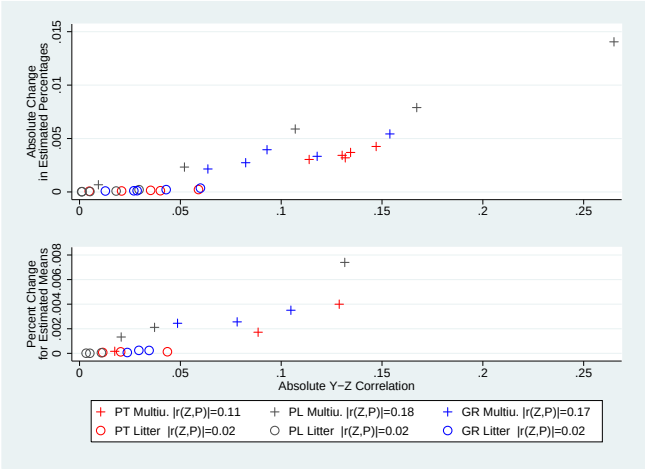

Figure 5: Change in estimated percentage and means after nonresponse adjustment

4.1.1 Changes after weighting

Weighting survey variables to adjust for nonresponse can have an effect on the point estimate as well as the estimated variance of a survey statistic. For now we focus on effects on the point estimate. In this illustrative example we will use two of the interviewer observations, multi-unit housing and presence of litter. Those two variables represent the range of z-p correlations and z-y correlations we have seen earlier. We will show the effect on the point estimate separately for the three countries.

The difference between the adjusted and unadjusted point estimates is displayed in Figure 5. The upper panel shows the absolute change in estimated percentages for five dichotomized key survey variables[10]. The change in estimated percentages is computed for each of the three countries (indicated in different colors) and with either one of the two interviewer observations used as nonresponse adjustment cell. As a reminder, the correlation between the interviewer observation and the response propensity is written into the respective legend. The lower panel follows a similar structure for the mean of three survey variables ("TV watching", "Trust in Legal System" and "Trust in the UN"). This time the y-axis is percentage change for estimated means. The graph illustrates nicely what we discussed earlier and highlights a couple of additional points:

- First, one can see the positive relationship between the correlation of Z and Y variables (y-axis in the graph) and the change in the estimated means or percentages. As expected by the work of Little and Vartivarian (2005), the interviewer observations affect the survey point estimates only when there is a correlation between the survey variables and the key survey statistics.

- Second, the estimated change is lower for interviewer observations (Z's) that have a lower correlation with response propensity; here the observation of litter in the neighborhood.

- Third, the changes in point estimates are very small. Here the greatest change in the point estimate was found for Poland for the item "Safety walking after dark". The point estimate for this fear of crime indicator, using just the selection weights, was 69.1%. After using the interviewer observation of housing structure as adjustment variable, the new point estimate is 67.7%. The different of 1.4% percentage points is displayed in the graph. Here the correlation between the fear of crime indicator and the interviewer observation, multi-unit housing structure, had an absolute correlation of $|0.27|$ and the correlation of the interviewer observation and the response propensity was $|0.18|$.

- Fourth, in our example analysis, the patterns are very similar across the three countries. Exploratory analyzes for other countries showed that this is not necessarily the case. But the high rates of missing data prevent us from drawing further inferences about this matter.

5 Discussion

The purpose of this paper was to examine interviewer observation data from the ESS to check the extent to which they could be useful for nonresponse adjustments. For

[10]Four of the five variables were originally asked with Likert type scales. Those were dichotomized to ensure an even split. In the case of 4 category variable ("Interest in politics" and "Safety of walking after dark", it meant grouping the two top categories together and the two bottom categories together. In cases where 5 categories were present (as for the political attitude items and the self reported health item), the middle category was added to either the top or bottom category depending on the specific distribution of the variable. The fifth variable "Not a victim of a crime" was already a dichotomized variable and did not need to be transformed.

the ESS data we are missing external criteria to evaluate the effectiveness of nonresponse weights created on the basis of interviewer observations. In this preliminary analysis we can only discuss the potential of the interviewer observations (based on the correlations with the observed response and the answers given by respondents on the variable of interest) to move the estimates. We can not evaluate if the weighted estimate is close to the 'true' value.

So far we have seen only weak correlations between the response propensity, the survey outcome, and the interviewer observations under study. All of these should be highly correlated to see improvements in the means square error (MSE) after nonresponse adjustement. It is possible that high interviewer variability is driving the attenuated correlations observed between (Y, Z) and between (P, Z). In the absence of an interpenetrated assignment of interviewer to sampling points, the ESS data do not allow systematic examination of this issues.

There are other data sets available that might shed light on this measurement problem. One of them is the Los Angeles and Neighborhood Survey (L.A. FANS). Unfortunately we did not had access to these data before giving this talk. Briefly, the L.A. FANS is a longitudinal study of 3,085 families in Los Angeles County and of the neighborhoods in which they live. Wave 1 of the survey included the collection of a household survey, a neighborhood survey, and a file of neighborhood services and characteristics based on census data and other administrative records. The neighborhood survey is a form completed by the field interviewers with the purpose of recording the physical and social characteristics of blocks in a systematic manner. In Wave 1, interviewer observations were completed in 422 sampled blocks with a mean of 2.3 independent observations completed for each block.

The availability of multiple assessments per block will allows us to compute estimates of interviewer variability across neighborhoods. An indication of how reliable these measures are can help interpret better the correlations observed for the ESS. It can also help to understand whether or not interviewers are able to make accurate and reliable observations of some neighborhood features. To our knowledge, no assessment of measurement error on interviewer observation has ever been done on the ESS or any other survey.

L.A. FANS can also help to alleviate much of the concerns associated with the application of the multinational ESS: (i) L.A. FANS has less than 1% of missing data for the interviewer observations - low risk of nonresponse bias - whereas ESS exhibits a variety of nonresponse patterns for these items, and (ii) L.A. FANS used a single training protocol for data collection - low risk of variability - whereas ESS countries present a variety of protocols developed by each participating country. Next year at the Joint Statistical Meeting we will report results from the L.A. FANS study.

References

Abraham, K. G., A. Maitland, and S. M. Bianchi (2006). Nonresponse in the american time use survey - who is missing from the data and how much does it matter? *Public Opinion Quarterly 70*(5), 676–703.

Bates, N., J. Dahlhamer, and E. Singer (2006). Privacy concerns, too busy, or just not interested? exploring the relationship between doorstep concerns and nonresponse bias. In *Paper presented at the 17th International Workshop on Household Survey Nonresponse, Omaha, Nebraska*.

Copas, A. J. and V. T. Farewell (1998). Dealing with non-ignorable non-response by using an 'enthusiasm-to-respond' variable. *Journal Of The Royal Statistical Society Series A-Statistics In Society 161*, 385–396.

Couper, M. P., E. Singer, and R. A. Kulka (1998). Participation in the 1990 decennial census - politics, privacy, pressures. *American Politics Quarterly 26*(1), 59–80.

Curtin, R., S. Presser, and E. Singer (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly 69*, 87–98.

Czajka, J., S. Hirabayashi, R. Little, and D. Rubin (1987). Evaluation of a new procedure for estimating income and tax aggregates from advance data. In *Statistics of Income and Related Administrative Record Research: 1986-1987*, pp. 109–135. Internal Revenue Service.

Czajka, J., S. Hirabayashi, R. Little, and D. Rubin (1992). Projecting from advance data using propensity modeling: An application to income and tax statistics. *Journal of Business & Economic Statistics 10*, 117–131.

Dalenius, T. (1983). Some reflections on the problem of missing data. In W. Madow and I. Olkin (Eds.), *Incomplete Data in Sample Surveys*, pp. 411–413. New York: Academic Press.

De Leeuw, E. and W. DeHeer (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. Groves, D. Dillman, J. Eltinge, and R. Little (Eds.), *Survey Nonresponse*, pp. 41–54. New York: Wiley.

Dillman, D., J. Eltinge, R. Groves, and R. Little (2002). Survey nonresponse in design, data collection and analysis. In R. Groves, D. Dillman, J. Eltinge, and R. Little (Eds.), *Survey Nonresponse*, pp. 3–26. New York: Wiley.

Duhart, D., N. Bates, B. Williams, G. Diffendal, and P. Chiu (2001). Are late/difficult cases in demographic survey interviews worth the effort? a review of several federal surveys. In *Proceedings of the Federal Committee on Statistical Methodology Research Conference*.

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

Groves, R. and S. Heeringa (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society A 169*, 439–457.

Groves, R. and K. McGonagle (2001). A theory-guided interviewer training protocol regarding survey participation. *Journal of Official Statistics 17*, 249–266.

Groves, R. and E. Peytcheva (2006, April). The impact of non-response rates on non-response error: A meta-analysis. In *Paper presented at the European Conference on Quality in Survey Statistics*, Cardiff.

Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly 70*(5), 646–675.

Kalton, G. and I. Flores-Cervantes (2003). Weighting methods. *Journal of Official Statistics 19*, 81–97.

Kalton, G. and D. Maligalig (1991). A comparison of methods of weighting adjustment for nonresponse. In *Proceedings of the 1991 Annual Research Conference*, pp. 409–428. Washington, DC: U.S. Bureau of the Census.

Little, R. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review 54*, 139–157.

Little, R. and D. Rubin (2002). *Statistical Analysis with Missing Data*. Hoboken, New Jersey: Wiley.

Little, R. and S. Vartivarian (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology 31*, 161–168.

Lynn, P. (2003). Pedaksi: Methodology for collecting data about survey non-respondents. *Quality and Quantity 37*, 239–261.

Philippens, M. and J. Billiet (2007). Monitoring and evaluating non-response issues and fieldwork efforts in the european social survey: Results of the analysis of contact form data. ESS Methodological Research Papers.

Politz, A. and W. Simmons (1949). An attempt to get the 'not at homes' into the sample without callbacks. *Journal of the American Statistical Association 44*, 9–31.

Raessler, S. and R. Schnell (2004). A comparison of multiple imputation and other unit-nonresponse compensating techniques in fear of crime studies. Mansucript presented at the Joint Statistical Meeting, Toronto.

Riede, T. and D. Asef (2006). The impact of contact times on the measurement of employment and unemployment by use of a telephone survey. In *Paper presented at Q2006 European Conference on Quality in Survey Statistics, Cardiff, UK*.

Rogers, A., M. A. Murtaugh, S. Edwards, and M. L. Slattery (2004). Contacting controls: Are we working harder for similar response rates and does it make a difference? *American Journal of Epidemiology 160*, 85–90.

Sampson, R. J. and S. W. Raudenbush (1999). Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. *American Journal Of Sociology 105*(3), 603–651.

Schnell, R. (1993). Die Homogenität sozialer Kategorien als Voraussetzung für 'Repräsentativität' und Gewichtungsverfahren. *Zeitrschrift für Soziologie 22*, 16–32.

Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen*. Opladen: Leske & Budrich.

Schnell, R. and F. Kreuter (2000a). Das DEFECT-Projekt: Sampling-Errors und Nonsampling-Errors in komplexen Bevölkerungsstichproben. *ZUMA-Nachrichten 47*, 89–101.

Schnell, R. and F. Kreuter (2000b). Untersuchungen zur Ursache unterschiedlicher Ergebnisse sehr ähnlicher Viktimisierungssurveys. *Kölner Zeitschrift für Soziologie und Sozialpsychologie 1*, 96–117.