Knocking on Respondents' Doors - Interviewers and Unit Non-Response in a Large Wealth Survey

Julia Le Blanc[1], Tobias Schmidt[2]
[1] Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, julia.le-blanc@bundesbank.de
[2] Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Tobias.schmidt@bundesbank.de

Abstract
In this paper we analyse the impact of interviewers and regional characteristics on unit non-response and non-contact in a voluntary face-to-face survey of households. We contribute to the existing literature by studying area and interviewer effects as well as interactions of both within a unified framework. The data for our analysis comes from the new German survey "Panel on Household Finances" (PHF), which over-samples wealthy areas in Germany. Making use of the special sampling design of the PHF, we analyse differences between wealthy and other regions as contact and cooperation behaviour across these groups may differ. Results from multilevel logistic regressions show that dissimilarities between areas with respect to wealth explain only a small part of the variance of households' response and contact behaviour. Interviewer effects on the contrary explain large portions of the observed differences in contact and response outcomes.

Key Words: interviewer effects, wealth survey, contact protocols, unit non-response

1. Introduction

Achieving high response rates is one of the goals of all voluntary sample surveys. With response rates falling in most household surveys (see e.g. de Leeuw and de Heer, 2002; Curtin et al., 2005) non-response and non-contact have become topical issues for survey practitioners and statisticians (Singer, 2006) alike. While statisticians are concerned with methods to correct for the impact of non-response bias on survey quality, survey practitioners discuss ways to minimize non-response in the first place. Many important aspects of the data gathering process and methods to increase response have been analyzed, including the role of interviewers. Besides interviewers' age, gender, motivation, clothing, and attitudes, behavioural aspects have been shown to be significant factors for interview outcomes.

Our paper is related to this literature, and in particular to the studies on the role of face-to-face interviewers for survey non-response and non-contact. The main contribution of our study is to investigate how important interviewer effects are in determining contact and cooperation compared to effects caused by differences between wealthy and other regions[1]. We furthermore contribute to the existing literature by studying interaction effects between regional and interviewer effects.

The data we use for this analysis is from the first wave of a new German wealth survey (PHF - "Private Haushalte und ihre Finanzen" – Panel on Household Finances)[2]. This survey was conducted in 2010/11 by the survey company infas on behalf of Deutsche Bundesbank. The PHF contains detailed information on households' wealth socio-economic characteristics of households in Germany and is similar in content to the US Survey of Consumer Finances (SCF). Besides the PHF-data itself, detailed metadata on contact and response behaviour for over 10,000 households, as well as information on the demographics of interviewers and paradata on dwellings of sampled households is available. What makes the dataset unique is that the random sample is stratified by the wealth of an area (municipality or street section). This feature of the sample presents a solid basis for comparing contact and response patterns for wealthy and less wealthy regional units.

Our results from multilevel logistic regressions show that dissimilarities between wealthy and other areas explain only a small part of the variance of households' response behaviour. Interviewer effects on the contrary, explain large portions of the observed differences in contact and response outcomes. The interaction between interviewers and the regions they work in is non negligible.

Based on our results we conclude that survey managers should pay particular attention when assigning interviewers to areas with a high share of wealthy households. Furthermore our findings indicate that the over-sampling rate for wealthy households/regions does not have to account for differential response behaviour.

In the next section we review some of the literature that deals with interviewer and area effects in voluntary surveys. Section three describes our data in more detail and provides an overview of the variables we include in our analysis. The results of our analysis are reported in section four, before we draw conclusions and talk about recommendations for interviewer management in voluntary surveys in section five

2. Literature Review

In this section we review some of the exiting literature related to our analysis of interviewer and area effects. The most relevant studies for our analysis deal with interviewer effects on a very general level and effects that are related to the characteristics of an area a given sampled household lives in.

2.1 Interviewer effects

The role of interviewers in face-to-face and telephone interviews has received a lot of attention in the survey methodology literature (see Groves and Couper, 1998) for an overview). Both practitioners and researchers have been analysing interviewers'

[1] We will use the terms "area" and "region" interchangeably throughout the paper. Note, the "areas" we are looking at are not confined geographical areas, but are a combination of many units scattered all over Germany with the same characteristics, namely the same wealth level (see also section 2.2 below).

[2] See von Kalckreuth et al. (2012) and www.bundesbank.de/phf-research for additional information on the survey.

behaviour and devised strategies to increase their performance for a long time (Durbin and Stuart, 1951; Groves and McGonagle, 2001; Couper and Groves, 1992; O'Brien et al., 2002).

The importance of interviewers in face-to-face designs stems from their tasks at the centre of the survey process: to locate the address of the respondent and achieve contact with the target person or household, and to convince the respondent to participate in the survey (Durrant et al., 2010; Blohm et al., 2006; de Leeuw and de Heer, 2002).[3]

Even though survey designers try to standardize the data collection process as much as possible, interviewer effects still exist. The reason for this is that interviewers are very heterogeneous along several dimensions, e.g. with respect to demographics, experience, skills or attitudes. What is more, contacting target units and achieving cooperation require different skills and as O'Muircheartaigh and Campanelli (1999) write "… some interviewers [are] better at reducing the refusal side of non-response and others [are] better at reducing the non-contact side of non-response."(p. 438).

The research on interviewer effects started with the analysis of interviewer socio-demographics. The evidence on the relationship between these characteristics and response behaviour is mixed (see Campanelli and O'Muircheartaigh, 1999). What is more, interviewer age, gender, and education seem to explain only a small part of the variation in response probabilities of target units (Groves and Couper, 1998). In recent years, interviewer experience, skills, and respondent-interviewer interactions have received considerable attention among survey methodologists (Campanelli et al., 1997; Durrant et al., 2010; Blom et al., 2010). Very influential contributions in this area have been made by Couper, Groves and their co-authors (Couper and Groves, 1992; Groves and Couper, 1996; Groves and Couper, 1998; Groves and McGonagle, 2001). They develop a theoretical basis for the process of the interviewer–respondent interaction and argue that experienced interviewers adapt and tailor their contact and cooperation strategies for each household, while inexperienced interviewers tend to fail to maintain interaction and create soft refusals.

Most of the literature cited above has analysed the effect of interviewers and interviewer characteristics on non-response alone, i.e. not separating non-contact and non-response. Some more recent studies look separately at cooperation and contact outcomes.

Schräpler et al. (2010), a paper closely related to ours, separate contact and cooperation outcomes. In their paper, they look at interviewers and response behaviour in a random sample of households from the German Socio Economic Panel (GSOEP), using a multi-level modelling approach. The authors find in their two level models, that interviewer effects do exist for contact and cooperation outcomes, but that they are rather small.

Another example is Blohm et al. (2006). They analyse data from the large ALLBUS survey conducted in Germany in 2000 using multi-level modelling. Their main goal is to assert the influence of interviewers' contact behaviour on successful contact and

[3] Conducting the interview is an important task of the interviewer and the analysis of the impact of interviewers on data quality has a long tradition in survey methodology research, see e.g. Hansen and Marks (1958), Feldman et al. (1952), O'Muircheartaigh and Campanelli (1998), Schnell and Kreuter (2005). In this paper we will not address the impact of interviewers on data quality, but only on contact and cooperation outcomes.

cooperation. In order to achieve contact, the mode of first contact – telephone or in-person - does not seem to make a difference. Likewise, interviewer demographics are only of minor importance for contact rates. For cooperation Blohm et al. (2006) find a negative effect on cooperation for both being full-time employed in addition to working part time as an interviewer and interviewer age.

Lipps and Benson (2005) present evidence on the influence of interviewers' contact behaviour on non-response. They show that an in-person contact is potentially more likely to lead to a successful contact on the first contact attempt than a contact attempt via telephone. However, experienced interviewers seem to be able to moderate this effect.

In summary, the literature shows that interviewers and their behaviour are very heterogeneous and their influence on contact and cooperation outcomes in face-to-face surveys may be different.

2.2 Area Effects

Besides interviewers, the areas or geographical regions where the respondents live have been identified as a source of potential variation in response and contact rates (Hox and DeLeeuw, 2002).

In recent years information on dwellings and area of residence of potential respondents has been collected to supplement call-record data and data on a higher level of regional aggregation (Kreuter and Casas-Cordero, 2010). This data has been used, in combination with geo-referenced data, to describe target households and to analyse their response behaviour (e.g. Schräpler et al., 2010; Blom et al., 2010). It is not surprising that the structure of the area a household lives in influences contact and cooperation rates.

O'Muircheartaigh and Campanelli (1999) cite literature showing that non-contact and refusal increases with urbanicity. Campanelli et al. (1997) report effects of the household and age structure of a region on non-contact. It seems especially hard to reach young people and people living alone. Cooperation can be expected to be low in areas with a lot of renters, old persons, and migrants, according to the UK data analysed by Campanelli et al. (1997). Schräpler et al. (2010) also contribute to this discussion; they find that some micro-geographic indicators at the level of street sections do actually explain some of the variance in response and contact rates. However, they summarize their findings, by noting that “the effect sizes of these effects are negligible” (abstract). Studies on the geographic distribution of refusals in the Current Population Survey (CPS) in the US find significant regional differences in response behaviour (DeMaio, 1980). They also show that rural respondents are more likely to cooperate than urban respondents.

In this paper we are concerned with the differential response behaviour of households in wealthy and less wealthy areas. The existing empirical evidence on the response behaviour of wealthy households is mixed and scarce. Previous studies on non-response in wealth surveys have pointed out that response rates are lower for wealthier households as the wealthy tend to be less likely to respond, both because of outright refusals as well as higher non-contact rates (see Kennickell, 2007; Bover, 2004; Faiella, 2002).[4] For the French Enquete de Patrimoine from 2004, which over-samples white–collar workers and

[4] The higher non-response rates of the wealthy are one of the reasons usually cited for over-sampling them in sample surveys.

areas in and around Paris, the response rates do not differ between wealthy and less wealthy areas. Kennickell (1998) even finds for the SCF that the response rates among the wealthy are comparatively high. Campanelli et al. (1997) show that in the UK non-response and non-contact is especially high in areas with less well-off residents.

2.3 Interactions between Interviewer and Area Effects

Some studies address the issue of whether interviewer or area effects are more important. Blom et al. (2010) e.g. find that interviewer effects are more important than country effects for explaining differences in contact success. Using data from the European Social Survey (ESS), they show that interviewers account for approximately a quarter of observed variances in contact probabilities. For cooperation conditional on contact, interviewer effects are rather small in Blom et al. (2010), they explain only 8% of the variance. O'Muircheartaigh and Campanelli (1999) find that in the British Household Panel Survey (BHPS), the variance between different interviewers exceeds the variance between different areas. In another paper Campanelli and O'Muircheartaigh (1999) show that the interviewer effect for the refusal part of non-response is maintained at the household level but the role of the interviewer in the non-contact component of non-response disappears and the area effect is the important effect. Lipps and Benson (2005) find for data from the Survey of Health, Ageing and Retirement in Europe (SHARE) on eleven countries that interviewers explain more variation in contact probabilities than countries, and a large share of the variance of cooperation probabilities.

2.4 Hypothesis and Implications for the Empirical Analysis

The empirical analysis below combines the findings cited above. We follow the literature and argue that interviewer effects explain a larger part of the variation in response and contact rates than area effects. Even though the literature is not clear on “wealth effects”, we hypothesise that area effects should gain importance if wealth indicators are used to classify regions. To test this hypothesis we estimate a series of multi-level models including no explanatory variables.

Our approach with respect to area effects differs from the existing approaches in one important aspect. Rather than using a sample of geographically defined regions, we use the a priori stratification of the sample by wealth of a regional unit (municipality or street section) for our analysis. The regions in our analysis are therefore not geographically coherent or confined areas but consist of different smaller areas scattered across geographical regions within Germany that share the same categorisation according to our definition of wealth. To give an example, Frankfurt and Munich may belong to the “wealthy large cities” group, while they are geographically located in different Länder (Hessia and Bavaria), districts, and municipalities.

3. Data and Variables

The data for this paper is from the German wealth survey “PHF” (“Private Haushalte und Ihre Finanzen” – Panel on Household Finance) conducted in 2010/11 by infas Gmbh on behalf of the German Bundesbank. This survey is the German contribution to the Eurosystem Household Finance and Consumption Survey (HFCS), an international effort to collect harmonized data on household finances. For the German survey 20,501 households were drawn using a stratified random sample with an eye on over-sampling wealthy households based on micro-geographic indicators (see subsection below). The target was to realize face-to-face CAPI interviews with about 4,000 households. To support interviewing and contacting, the households received an advance letter from infas containing the name and phone number of the interviewer assigned to them. These letters also contained information on data protection, a letter from the president of the Deutsche Bundesbank and a brochure with additional information on the survey.

The PHF field phase consists of two major parts, an initial field phase and a "re-launch" field phase. The latter was initiated because the initial field phase yielded a low response rate. The second field phase includes some methodological changes, which make a comparison or joint analysis with the data from the initial field phase problematic[5]. In our analysis we will therefore only use data from the initial field phase. The initial PHF field phase lasted 23 weeks. It began in early September of 2010 and ended on February 28, 2011. Beginning in March 2011, a special conversion phase was initiated which lasted until July 15, 2011. During the conversion phase households that had not been reached were contacted centrally by the survey agency's CATI interviewers to make appointments. These contact attempts are not part of our analysis. The first contact attempt by a face-to-face interviewer was made on September 10, 2010 and the first interview conducted on September 25th, 2010. At the beginning of September, 212 trained interviewers were deployed to the field. Each interviewer was assigned to at least one sample point,[6] consisting of 45 addresses each, resulting in a gross sample of 10,260 addresses. The total gross sample was split into two parts, one with 8,208 addresses and one with 2,056 addresses.[7]

The response rates were low in comparison with similar studies, as only 1,765 interviews were completed in the initial field phase. Corrected for 739 ineligible addresses this means that only 19% of eligible respondents answered.[8]

What makes the dataset particularly interesting for our study is the availability of detailed metadata on contacts and information on the building types and neighbourhoods the households live in. As regards the first type of data, the information on contacts comes directly from software used by the survey agency to track field work in the PHF.[9] The tool collects data for each contact on the time and mode of contact (telephone, in person), the interviewer attempting the contact, the outcome of the contact and an indicator for the sample point the interviewer works in. In total almost 33,500 contact attempts were made by face-to-face interviewers. The quality of this data is very high as interviewers are given clear incentives to record each attempt carefully. They are reimbursed for their efforts only if the contact attempt is registered in this database.

---

[5] These included: shortening the field phase to 8 weeks, interviewers were no longer assigned to specific sample points by infas, but were allowed to choose sample points they want to work in, interviewers were required to contact each "undecided" household at least once every week, households in sample points with bad housing conditions were given an extra incentive.
[6] The addresses interviewers received were mainly located in the geographic vicinity of their place of living.
[7] The second tranche of addresses was given to the field managers at the beginning of December 2010, when several interviewers had completed the work on the addresses they had initially received.
[8] Some of the addresses given to the interviewers were not completed at the end of the field phase. We treat these cases as refusals or non-contact, respectively, for the calculation.
[9] The interviewers were required to contact the households during different times of the day and week, and to visit the household at least once in person. They were required to register each contact with the household. The recorded data was used to monitor the interviewers' behavior and to determine the compensation they would receive for their contact attempts.

We also use information from the infas geodata database, which contains micro-geographic information for street sections and urban districts. This information is available for all sampled addresses and can be linked directly to the respective address in the sample register.

3.1 Outcome and Response Variables

We follow the existing literature and analyse the impact of certain characteristics on cooperation and contact, the two key variables of interest for survey practitioners and researchers.

The contact variable takes the value one if the interviewer made contact with a person living at the household's address, and zero otherwise. The response codes leading to labeling a case as "contact established" include among others the interview itself as well as all types of refusals.[10] No distinction was made between in-person and telephone contacts initiated by face-to-face interviewers. Ineligible addresses and addresses without a single contact attempt are not considered.[11] Descriptive statistics show that about 42% of all contact attempts were successful. Of the 9,295 eligible households with at least one contact attempt 8,544 were contacted successfully. This means that 92% of eligible households were contacted.

Cooperation is constructed as a binary variable. It is one if the interviewer realized an interview with the household head ("financial knowledgeable person") and zero otherwise. The low cooperation rates were already mentioned above. In sum, 1,765 interviews were realized, 20.7% of all successfully contacted households.

3.2 Variables Defining Levels

In the empirical part of the paper we estimate multi-level logit models (see below) with three levels. Level one (the lowest level) is always the unit of observation, in our case an individual household. Each household is assigned to one interviewer. The interviewers form the second level. Each interviewer is assigned a unique identifier, which makes the modelling of this level straightforward. Our third (top level) level will be regional indicators related to an area's wealth. The basic specification mimics the stratification employed to realize the over-sampling. Small municipalities with less than 100,000 adult inhabitants, were classified according to the share of tax payers, whose taxable income exceeds a certain threshold. In large cities with 100,000 or more adult inhabitants, the stratification was done on the level of street sections. Street sections which are characterised by a high quality of buildings and a substantially above average purchasing

---

[10] The response codes for "contact established" are: Reference Person (RP) may be reached within the next few days, RP cannot be reached during the field phase, RP cannot be reached, because third person refuses, appointment for next contact scheduled, successful interview, interview stopped – will be continued, interview stopped – RP refuses to continue, refusal because of: lack of interest, topic, no time, personal contact not welcome, too many surveys, data protection, length of interview, illness, not allowed to participate, language, other reasons, confidentiality.
RP not speaking any survey language or RP asking for Turkish, Russian, English or Polish version of the questionnaire

[11] The population considered for the PHF includes only private non-institutionalized households, other households were considered ineligible. Addresses are also considered ineligible if the person drawn has moved to an unknown location or is unknown at the given address.

power per resident are considered wealthy.[12] This design results in four different types of "regions" at the top level: wealthy municipalities with less than 100,000 adult inhabitants, other municipalities with less than 100,000 adult inhabitants, "wealthy" street sections in large municipalities with more than 100,000 adult inhabitants and other street sections in large municipalities. Table 1 gives some descriptive statistics for these four groups.

[INSERT TABLE 1 ABOUT HERE]

For extensions of the basic specification and robustness checks we make use of several other indicators related to the wealth of a region (see below). The four other indicators are: 1) quality of residential area for street sections; 2) a social class index for street sections, 3) the purchasing power per resident, and 4) social class indicators at the urban district level. All these indicators are available from the infas geodata database for almost our whole gross sample.[13] The quality indicators of residential area takes on four values: bad or very bad, satisfactory, good, prime or very good residential area. Infas geodata provides the number of households in each street section that fall into one of these categories. A street section is assigned to the category to which the highest number of households living in that district belongs. The social class index for street sections is a combination of several other indicators included in the infas geodata database and ranges from 0 to 200. In our analysis we will create four groups, based on quartiles calculated from the index.

At the urban district level we make use of the purchasing power per resident and again assign urban districts to four groups based on quartiles. The classification of households by social class is used to construct another set of variables identifying the third level of the multi-level model. Here again we use the procedure outlined above and assign an urban district to the class (upper, upper middle, middle, lower middle and lower) which contains the highest number of households.

4. Estimation Method

We investigate how interviewer and area characteristics affect contact (i.e. whether contact with the household was made or not) and cooperation (i.e. whether the sample unit was interviewed or not, after contact was made). Since our survey data have a hierarchical structure due to multiple nesting, we use a multilevel logistic regression model as is now standard in the literature (e.g. Blom et al., 2010; Schräpler et al., 2010; Durrant et al., 2010; Blohm et al., 2006).[14] The advantage of using multilevel models over conventional estimation procedures like logit estimation is twofold: First, disregarding the nested or clustered structure of the data can result in biased estimates (Hox, 1998). Second, the structure of the model allows investigating between-cluster variances and analysis at different levels of the hierarchy of the data (Carle, 2009).

The simplest forms of multilevel models for response behaviour in surveys use only two levels, the level of the household and the interviewer. Three and higher-level models

[12] A more detailed description of the sample design is available in German ("Methodenbericht") on the the PHF's website (www.bundesbank.de/phf-research).
[13] There is a very small number of street sections (in small municipalities) that could not be linked to the infas geodata database.
[14] The books by Hox (2010) and Rabe-Hesketh and Skrondal (2008) are excellent introductions to multi-level modelling.

have also been investigated, with the additional level being the country (e.g. Blom et al., 2010) or area (Campanelli and O'Muircheartaigh, 1999). Usually the structure is strictly hierarchical, i.e. one household is assigned to only one interviewer, and one interviewer works in only one country or area. This is not the case in the PHF. In the PHF the structure is not strictly hierarchical, because some interviewers (level 2) work in several regions (level 1).[15] In order to consider this additional complexity, we use a three-level cross-classified multilevel logistic model, with regions and interviewers cross-classified and households at the lowest level (level 3). The interpenetrated structure of the survey data reduces the problem of confounding between interviewer and "area" effects (O'Muircheartaigh and Campanelli, 1999). An additional advantage of this approach is that we can analyse the effects of interviewer-region pairs on contact and cooperation outcomes.

We estimate our models using the xtmelogit command implemented in STATA.[16] This module uses Gaussian quadrature to approximate the log likelihood function. The estimation of the log likelihood is a precondition for a variety of tests, like the LR-test for comparing nested models (see Hox, 2010).

Where necessary, we rescale the variances following the method described in Hox (2010) in order to make the variance components comparable across models. This method builds on McKelvey and Zavoina (1975) and Snijders and Boskers (1999) and uses information on the variance of the fixed part of the specification as ingredients for rescale factors for both the variances and the estimated coefficients of differently specified models.

There has been some discussion of using weights in multilevel analysis (Rabe-Hesketh and Skrondal, 2006; Grilli and Pratesi, 2004; Carle, 2009), but no clear guidance on how to handle weights in cross-classified data exists. Carle (2009) even states that "… no work has examined handling design weights in this [cross-classified] situation". We therefore do not attempt to conduct weighted multi-level analysis, even though over-sampling was employed in the sample procedure for our survey data.

5. Results

5.1 Descriptive Evidence

The performance of interviewers in the PHF varies considerably. Only 193 interviewers completed at least one interview. The three most successful interviewers had 66, 51 and 50 completed interviews, respectively. The median is at 6 interviews per interviewer. A similar picture emerges for contact attempts: 18% of interviewers had on average less than two contact attempts per address, 50% had three or more per assigned address. Even though these figures have to be interpreted with care, since more successful interviewers may have fewer attempts, because they realize interviewers quicker/earlier, they underscore the substantial heterogeneity among interviewers. The response and contact rates for different areas also differ (see table 1), but to a much lesser degree than for different interviewers.

[15] In total 122 interviewers (58%) were assigned addresses in more than one PSU strata.
[16] See StataCorp LP (2009) for details on the program.

5.2 Econometric Evidence

Our main results are presented in Table 2. Interviewers are responsible for a larger share of the variance of contact and cooperation outcomes than the type of area the households live in. This difference is more pronounced for contact than for cooperation behaviour. As expected households themselves are mostly responsible for the observed differences. Their influence is much stronger for cooperation (91.5% of variance explained by households) than for contact (73-74%), however. Due to the interpenetrated design of the interviewer assignments (most interviewers work in at least two different types of strata), we also obtain results for an interaction between the two levels, interviewers and wealth areas. Of the total variance explained by interviewer and area effects, the interaction term accounts for 27% of contact and 37% of cooperation outcomes, respectively. While area effects alone are negligible, the indirect effect through the interviewer performance in a particular region is not.

[INSERT TABLE 2 ABOUT HERE]

Our classification of regions consists of two components: the size of a municipality and its wealth. In a hypothetical case where the wealthy and less wealthy households do not differ in their response behaviour, we may still observe variance between the regions if response behaviour differs with municipality size. In this case, our results would mirror the different sizes but not the wealth differences. However, including a dummy for municipalities with 100,000 or more inhabitants, has basically no effects on the results, only the area effect decreases marginally (see Table 2).

We conduct additional checks that deal with the grouping of regions by wealth and size[17]. We first ignore the municipality size and only retain the separation of wealthy and other municipalities. The qualitative results with respect to the difference between interviewer and area effects do not change. The interaction effect is lower, of course, as now there are only two areas. We then ignore the wealth differences between the regions and group them according to municipality size only. The interviewer effect still remains the dominant effect, however, the interaction term for contact is much more pronounced than in the previous estimations. We conclude from this that for contact outcomes the main difference is between interviewer performance in large and small regions. An additional robustness check splits the sample in one where interviewers and households are cross-classified and one where this is not the case. We first ignore the cross-level interaction and restrict the analysis to interviewers working in one area only. The interaction term is absorbed by the increased variance in interviewer effects, while the area effects remain low. For the sample with cross-classification we obtain results that are almost the same as those for the full sample. To sum up, the robustness checks confirm the results of the main specification – interviewer effects are more important than area effects, and area effects are by and large negligible.

A natural question to ask is whether our regional wealth indicators can actually differentiate between wealthy and less wealthy households. One of the main drawbacks of our stratification method is the use of micro-geographical indicators on wealth as opposed to direct wealth information for each household in our sample. However, we are confident that our stratification successfully allocates wealthy and less wealthy households into the respective categories. This is confirmed by preliminary data on

[17] Due to space limitations the results tables for these checks are not reported. They are available upon request.

responding households. Median income and median wealth are indeed significantly higher in our wealthy regions.

Given the richness of micro-geographic information, other classification rules can be tested. To analyze whether our results depend on the specific classification chosen and whether they would change if we used different indicators, we conduct a quasi-experiment. To be more precise, we take the wealth holdings of responding households and correlate those with micro-geographic indicators in order to identify the "optimal" indicators for stratification by wealth. The highest correlations are with the four indicators described in section 3 above (results are available upon request).

Table 3 shows that regardless of the classification indicator used for wealthy areas, interviewer effects dominate area effects with the interaction terms in between. Area effects never account for more than 1% of the total variance for both contact and cooperation outcomes. We conclude from this experiment that the results of the basic specification – interviewer effects are more important than effects of wealthy areas - are not specific to the chosen classification scheme for wealthy regions.

[INSERT TABLE 3 ABOUT HERE]

6. Conclusions and Future Research

In this paper, we analyze the relative importance of interviewer and area effects on contact outcome. We have shown that interviewers are the main source of heterogeneity in contact and cooperation outcomes. Whether a household lives in wealthy or less wealthy areas is only of minor importance in this respect.

Our study goes beyond others in analyzing an interaction effect between different types of areas and interviewers. We show that the interaction component of the variance is indeed important. To be more precise, we show that approximately 30% of the overall effect induced by interviewers is attributable to the combination of interviewer and area/wealth characteristics.

Our findings have implications for survey managers: First, special attention should be given to the assignment of interviewers to wealthy areas. Second, the over-sampling rate does not have to account for differential response behaviour in wealthy areas.

Achieving a high response rate is, of course, not the only goal of a survey aiming at high data quality. Further analysis is therefore necessary to assess the impact of stratifying regions by wealth on measures like item non-response, interview length, and estimation bias.

Tables and Figures

Table 1: Descriptive statistics for four types of regions used in the baseline specification

Wealth Strata	Gross Sample	Net Sample (# hhs)	Number of contact attempts	Number of successfully contacted hhs	Number of realized interviews
Small & Other	3,105	2,851	9,745	2,646	621
Small & Wealthy	3,015	2,745	8,964	2,555	517
Large & Other	2,300	2,045	7,678	1,817	315
Large & Wealthy	1,840	1,654	5,809	1,526	312
Total	10,260	9,295	32,196	8,544	1,765

Source: Authors' own calculations based on response data for the PHF provided by INFAS

Table 2: Results for three level cross-level logit models with the baseline classification of regions (large &wealthy, small & wealthy, large & less wealthy, small & less wealthy)

		Contact		Cooperation (conditional on contact)	
fixed part	constant	included	included	included	included
	Dummy large muni.		included		included
random part	var(R.area)	0.057	0.023	0.006	0.000
	var(inter)	0.812	0.809	0.186	0.186
	var(areaXinter)	0.320	0.311	0.114	0.116
	var(fixed)	3.290	3.290	3.290	3.290
ICC (Share in total variance)					
	var(inter)	18.1%	18.3%	5.2%	5.2%
	var (R.area)	1.3%	0.5%	0.2%	0.0%
	var(areaXinter)	7.1%	7.0%	3.2%	3.2%
	var(R.Area)+var(area)+var(areaxint)	26.5%	25.8%	8.5%	8.4%
	Variance explained by households	73.5%	74.2%	91.5%	91.6%
	var(areaxinter)+var(inter)	25.3%	25.3%	8.3%	8.4%
	var(areaxinter)+var(R.area)	8.4%	7.5%	3.3%	3.2%
Integration points		3	3	3	3
Deviance		4827.68	4824.91	8484.62	8480.60
Obs		9,295	9,295	8,544	8,544

Source: Authors' own calculations

Table 3: Results for three level logit models with different types of classification

		Contact	Cooperation (cond.)	Contact	Cooperation (cond.)	Contact	Cooperation (cond.)	Contact	Cooperation (cond.)
Regional Classification		Street section: quality of residential area		Street section: social class index		Urban district: purchasing power per resident [€]		Urban district: social classes	
fixed part	constant	included	included	included	included	included	included	included	included
random part	var(R.area)	0.046	0.007	0.069	0.002	0.069	0.002	0.040	0.001
	var(inter)	0.919	0.269	0.972	0.250	0.972	0.250	0.863	0.244
	var(areaXinter)	0.222	0.000	0.129	0.050	0.129	0.050	0.225	0.041
	var(fixed)	3.290	3.290	3.290	3.290	3.290	3.290	3.290	3.290
ICC (Share in total variance)									
	var(inter)	20.5%	7.5%	19.8%	7.3%	21.8%	7.0%	19.5%	6.8%
	var (R.area)	1.0%	0.2%	0.9%	0.2%	1.5%	0.1%	0.9%	0.0%
	var(areaXinter)	5.0%	0.0%	5.3%	0.4%	2.9%	1.4%	5.1%	1.1%
	var(R.Area)+var(area)+var(areaxint)	26.5%	7.7%	25.9%	7.9%	26.2%	8.4%	25.5%	8.0%
	Variance explained by households	73.5%	92.3%	74.1%	92.1%	73.8%	91.6%	74.5%	92.0%
	var(areaxinter)+var(inter)	25.5%	7.5%	25.1%	7.7%	24.7%	8.4%	24.6%	8.0%
	var(areaxinter)+var(R.area)	6.0%	0.2%	6.2%	0.6%	4.4%	1.4%	6.0%	1.2%
# inter working in more than one area		200		201		125		171	
Integration points		3	3	3	3	3	3	3	3
Deviance		4794.62	7595.06	6699.55	7608.73	6692.07	7609.55	4796.30	8443.29
Obs		9,233	8,024	9,281	8,040	9,281	8,040	9,250	8,504

Acknowledgements

This paper represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or its staff. We would like to thank Iris Noack for outstanding research assistance.

References

Blohm, M., J. J. Hox, and A. Koch (2006), The Influence of Interviewers' Contact Behavior on Contact and Cooperation Rate in Face-to-Face Household Surveys, International Journal of Public Opinion Research 19 (1), 97-111.

Blom, A. G., E. De Leeuw, and J. J. Hox (2010), Interviewer Effects on Nonresponse in the European Social Survey, MEA Discussion Papers 202-2010, Mannheim.

Bover, O. (2004), The Spanish Survey of Household Finances (EFF): Description and Methods of the 2002 Wave, Banco de Espana Occasional Paper No. 0409.

Campanelli, P. and C. O'Muircheartaigh (1999), Interviewers, Interviewer Continuity, and Panel Survey Nonresponse, Quality & Quantity 33, 59-76.

Campanelli, P., P. Sturgis, and S. Purdon (1997), Can You Hear Me Knocking: an Investigation into the Impact of Interviewers on Survey Response Rates, National Centre for Social Research, London, UK.

Carle, A. C. (2009), Fitting Multilevel Models in Complex Survey Data with Design Weights: Recommendations, BMC Medical Research Methodology 9 (49), 1-13.

Couper, M. P. and R. M. Groves (1992), The Role of the Interviewer in Survey Participation, Survey Methodology 18, 263-277.

Curtin, R., S. Presser, and E. Singer (2005), Changes in Telephone Survey Nonresponse over the Past Quarter Century, Public Opinion Quarterly 69, 87-98.

de Leeuw, E. and W. de Heer (2002), Trends in Household Non-Response: A Longitudinal and International Comparison, in: R. M. Groves, D. A. Dillman, J. L. Eltinge and R. J. A. Little (eds), Survey Non-Response, Wiley, New York, 41-54.

DeMaio, T. (1980), Refusals: Who, Where and Why?, Public Opinion Quarterly 44 (2), 223-233.

Durbin, J. and A. Stuart (1951), Differences in Response Rates or Experienced and Inexperienced Interviewers, Journal of the Royal Statistical Society A 114 (1), 163-206.

Durrant, G. B., R. M. Groves, L. Staetsky, and F. Steele (2010), Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys, Public Opinion Quarterly 74 (1), 1-36.

Faiella, G. D. A. a. I. (2002), Non-response behaviour in the Bank of Italy's Survey of Household Income and Wealth, Banca d'Italia Temi di discussione No. 462.

Feldman, J. J., H. Hyman, and C. W. Hart (1952), A Field Study of Interviewer Effect on the Quality of Survey Data, Public Opinion Quarterly 15 (4), 734-761.

Grilli, L. and M. Pratesi (2004), Weighted Estimation in Multilevel Ordinal and Binary Models in the Presence of Informative Sample Designs, Survey Methodology 30 (1), 93-103.

Groves, R. M. and M. P. Couper (1996), Contact-Level Influences on Cooperation in Face-to-Face Surveys, Journal of Official Statistics 12 (1), 63-83.

Groves, R. M. and M. P. Couper (1998), Nonresponse in household interview surveys, Wiley, New York.

Groves, R. M. and K. McGonagle (2001), A Theory-Guided Interview Protocol Regarding Survey Participation, Journal of Official Statistics 17, 249-265.

Hansen, R. H. and E. S. Marks (1958), Influence of the Interviewer on the Accuracy of Survey Results, Journal of the American Statistical Association 53 (283), 653-655.
Hox, J. J. (1998), Multilevel Modeling: When and Why, in: I. Balderjahn, R. Mathar and M. Schader (eds), Classification, Data Analysis and Data Highways, Springer, New York, 147-154.
Hox, J. J. (2010), Multilevel Analysis: Techniques and Applications (Quantitative Methodology), Taylor & Francis, London.
Hox, J. J. and F. DeLeeuw (2002), The Influence of Interviewers' Attitude and Behaviour on Household Survey Nonresponse: An International Comparison, in: R. M. Groves, D. A. Dillman, J. L. Eltinge and R. J. A. Little (eds), Survey Nonresponse, Wiley & Sons, New York, 103-120.
Von Kalckreuth , U., M. Eisele, J. Le Blanc, T. Schmidt, J. Zhu (2012) The PHF: a comprehensive panel survey on household finances and wealth in Germany. Bundesbank Discussion Paper 13/2012, Frankfurt am Main.
Kennickell, A. B. (1998), Analysis of Nonresponse Effects in the 1995 Survey of Consumer Finances, Federal Reserve Board Paper Presented at the Joint Statistical Meeting of 1997, Washington, D.C.
Kennickell, A. B. (2007), The Role of Over-Sampling of the Wealthy in the Survey of Consumer Finances.
Kreuter, F. and C. Casas-Cordero (2010), Paradata, RatSWD RatSWD Working Paper Series 136, Berlin.
Lee, E. S., R. N. Forthofer, and R. J. Lorimor (1989), Analysing Complex Survey Data, Sage, Newbury Park.
Lipps, O. and G. Benson (2005), Cross-National Contact Strategies, Proceedings of the Survey Research Section of the American Statistical Association, American Statistical Association, Alexandria, VA.
McKelvey, R. and W. Zavoina (1975), A Statistical Model for the Analysis of Ordinal Level Dependent Variables, Journal of Mathematical Sociology 4, 102-120.
O'Brien, E. M., S. Mayer, R. M. Groves, and G. E. O'Neill. 2002, "Interviewer Training to Increase Survey Participation." Proceedings of the Survey Research Methods Section, American Statistical Association (2002): 2502-2507: New York.
O'Muircheartaigh, C. and P. Campanelli (1998), The Relative Impact of Interviewer Effects and Sample Design Effects on Survey Precision, Journal of the Royal Statistical Society A 161, 63-77.
O'Muircheartaigh, C. and P. Campanelli (1999), A Multilevel Exploration of the Role of Interviewers in Survey Non-Response, Journal of the Royal Statistical Society A 162 (3), 437-446.
Rabe-Hesketh, S. and A. Skrondal (2006), Multilevel modelling of complex survey data, Journal of the Royal Statistical Society A 169 (4), 805-827.
Rabe-Hesketh, S. and A. Skrondal (2008), Multilevel and Longitudinal Modeling Using Stata, Stata Press, College Station, Texas.
Schnell, R. and F. Kreuter (2005), Separating Interviewer and Sampling-Point Effects, Journal of Official Statistics 21 (3), 389-410.
Schräpler, J.-P., J. Schupp, and G. G. Wagner (2010), Individual and Neighbourhood Determinants of Survey Nonresponse, SOEPpapers 288, Berlin.
Singer, E. (2006), Nonresponse bias in Household Surveys, Public Opinion Quarterly 70 (5), 637-645.
Snijders, T. A. B. and R. Boskers (1999), Mulitlevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling, Sage, Thousand Oaks, CA.
StataCorp LP (2009), STATA Longitudinal Data/Panel Data Reference Manual - Release 11, Stata Press, College Station, TX.