The German Socio-Economic Panel Study

260

SOEP Survey Papers

Series C – Data Documentations

SOEP — The German Socio-Economic Panel Study at DIW Berlin

2014

Sampling, Nonresponse, and Weighting in the 2011 and 2012 Refreshment Samples J and K of the Socio-Economic Panel

Martin Kroh, Konstantin Käppner, Simon Kühne

DIW SOEP

Running since 1984, the German Socio-Economic Panel Study (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing.

The SOEP Survey Papers is comprised of the following series:

Series A – Survey Instruments (Erhebungsinstrumente)

Series B – Survey Reports (Methodenberichte)

Series C – Data Documentations (Datendokumentationen)

Series D – Variable Descriptions and Coding

Series E – SOEPmonitors

Series F – SOEP Newsletters

Series G – General Issues and Teaching Materials

The SOEP Survey Papers are available at
http://www.diw.de/soepsurveypapers

Editors:
Prof. Dr. Gert G. Wagner, DIW Berlin and Technische Universität Berlin
Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:

Martin Kroh, Konstantin Käppner, Simon Kühne. 2014. Sampling, Nonresponse, and Weighting in the 2011 and 2012 Refreshment Samples J and K of the Socio-Economic Panel. SOEP Survey Papers 260: Series C. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)

Contact: DIW Berlin
SOEP
Mohrenstr. 58
10117 Berlin

Email: soeppapers@diw.de

MARTIN KROH, KONSTANTIN KÄPPNER, SIMON KÜHNE

SAMPLING, NONRESPONSE, AND WEIGHTING IN THE 2011 AND 2012 REFRESHMENT SAMPLES J AND K OF THE SOCIO-ECONOMIC PANEL

München, 2014

Contents

List of Tables

List of Figures

1 Introduction

For prospective panel surveys, the implementation and integration of refreshment samples after wave 1 serves several purposes. First, the secular loss in sample size by cumulative nonresponse reduces the efficiency of sample-based estimates in later waves. Moreover, the longitudinal accumulation of weighting-based corrections for selective attrition rates will inflate design effects and thus reduce efficiency of the sample (Schonlau et al. 2013). Since retention rates in household panel surveys are fairly high and hover after the first two waves at more than 90 percent in most cases (Kroh 2014), a substantial loss in efficiency will take effect only after several waves. However, the long-term target of the Socio-Ecomnomic Panel (SOEP) to cover decades of social change, eventually reduces sample size. In this situation, replacement of non-responding households by randomly selected households from the same population represents a possible solution.

Changes in the underlying population are the second reason for new samples in prospective panel studies, such as SOEP. New immigrant households, for instance, who arrive after the sampling of the original members of the panel are by definition excluded from the study. So called enlargement samples cover these new cases to the target population after wave 1 of the panel. The Socio-Economic Panel Study (SOEP) has a tradition of regular refreshment as well as enlargement samples. Enlargement samples covering changes in the target population, i.e. private households in Germany, are Sample C in 1990 (households in East Germany), Sample D in 1994/5 (households migrating to West-Germany after the initial sampling in 1984), and Sample M in 2013 (households migrating to Germany since 1995 and households including children of immigrants). Cross-sectional refreshment samples compensating for panel attrition in the existing samples (SOEP 2012: 6) are Sample E (in 1998), F (in 2000), H (in 2006), and the present Samples J (in 2011) and K (in 2012). Refreshment samples of special populations, sometimes referred to as boost samples, were implemented in 2002 (sample G, high-income households) and 2009 (sample L, large families, single parents, and low-income families).

Like any other survey, the gross-samples of J and K are affected by nonresponse. The problem of units of analysis such as persons and households not participating in surveys – better known as "unit nonresponse" – is one of the major challenges faced by researchers when aiming for inference from a survey sample to a population. Moreover, due to rising rates of refusal and non-contact reported by researchers (e.g. Curtin et al. 2005), the challenge is growing over time. Depending on the mechanisms governing it, unit nonresponse can lead to bias in samples[1] and as a consequence in scientific findings in general (Bethlehem et al. 2011: 21). This bias is prevalent in estimates of means, effect coefficients, and other parameters of interest. Thus, nonresponse-bias may pose a threat to scientific inferences. It is in researchers' best interest to reduce nonresponse a priori, to understand and to document it ex-post, as well as to find ways to account for it in further use of survey data for statistical estimation.

While the first purpose of this research note is to document the sampling procedure of the latest two general Refreshment Samples J and K,[2] the second purpose is to document our approach to account for nonresponse in wave 1 of these samples. Its aim is to analyze participation and non-participation of households in the first waves of the 2011 and 2012 Refreshment Samples J and K of the German Socio-Economic Panel. Drawing upon the existing techniques developed to correct for nonresponse, the results will then be used to generate nonresponse-weights, that themselves can be used to account for nonresponse in substantive analyses of the data in form of weighting variables. Using the combination of design and nonresponse-weights, researchers may make more valid inferences from Samples J and K and may enhance the explanatory power of their research (Lumley 2010: 136). A major obstacle of any nonresponse study is to obtain information on those units of analysis who elect not to participate. This study draws on interviewer reports on the sampled addresses and geocoded information on the neighborhood, municipality, and county. To be clear, the study aims at balancing the gross sample and net sample with respect to a large number of household characteristics, but we do not interpret the correlates of non/response as reflecting causal relationships.

[1] For a discussion on when nonresponse leads to bias, see for instance Groves (2006).

[2] For information on the sampling procedures of the other SOEP subsamples see `www.diw.de/documents/dokumentenarchiv/17/diw_01.c.38951.de/dtc.409713.pdf` and sample-specific data documentations listed on the SOEP web pages `http://www.diw.de/de/diw_02.c.222858.de/dokumente.html`.

1.1 (Self)Selection into Surveys

Selectivity in the observed data may either be introduced by design, i.e. intended over-sampling of specific groups in the target population, or by the choice of the selected unit of analysis to participate in the survey or not. While design weights compensate for choices made by researchers in the sampling process and are therefore known, estimated nonresponse weights capture observable differences between the selected gross sample and the realized net sample (i.e., model-based weighting) on the one hand and between marginal distributions of the net sample and respective known marginal distributions of the underlying target population (i.e., post-stratification, raking, GREG) on the other hand.

Several theoretical explanations for unit nonresponse have been put forward to account for selective participation rates. However, they mostly come down to the explanation of an individual's decision to opt for or against participation (Groves et al. 1992: 475). Following the rational-choice paradigm this decision can be regarded as a result of cognitive evaluation of costs and benefits of participation. However, the case is more complicated when dealing with nonresponse in general population surveys such as th SOEP. Unlike other survey situations (e.g. clinical trials), due to non-participation itself very little is known about non-participants (Giraldo/Zuanna 2006: 296)[3] and evaluations of costs and benefits are not directly measurable in non-participants. Therefore, the aim is to identify other variables from other sources influencing the individual's perceptions of costs and benefits of participation. The selection of variables used in this paper's nonresponse analysis stems from the existing literature on nonresponse. Besides information on sampled households provided by the interviewer (e.g. Olson 2006; Keeter et al. 2006; Abraham et al. 2006), we also draw on geocoded information on the regional context of sampled households (e.g. Johnson et al. 2006). Following previous research on neighborhood effects in nonresponse, we consider, for instance, indicators related to affluence of regions and indicators for the level of social embeddedness[4].

[3] For instance, all of the factors mentioned by Groves et al. (1992: 480f) as being part of a generic unit nonresponse theory are not available for analysis here.

[4] See for instance the concept of "disadvantaged areas" in Johnson et al. (2006) or the concept of "isolation" of individuals at Durrant/Steele (2009).

1.2 Correction for (Self)Selection

Selective sampling and selective unit nonresponse as a source of bias can be dealt with by ex-post weighting of observed units. Propensity score weighting techniques assign observed units of analysis – in the present analysis: households – with more "importance" if they hold characteristics that are associated with lower selection probabilities and higher nonresponse.[5] More specifically, the weights are calculated as inverse observational probabilities. These consist of the known sampling probabilities and the estimated response probabilities conditional on sampling. An assumption about the nature of nonresponse needs to be made, however. The MAR ("missing at random") is most frequently used. It states that, unlike under the "missing completely at random assumption", participating units of analysis and those not participating differ only in observable characteristics. Therefore, some groups with specific (combinations of) characteristics opt for participation more or less frequently. However, when those differences in observed characteristics are controlled for, no systematic difference between participants and non-participants within groups exists (Schafer 1997: 10f). Thus, using weighted observed respondents, estimation of the parameters of interest still can provide valid inferences. In this study, the unknown probabilities are estimated using logistic regression and are then transformed into propensity weights (Kim/Kim 2007: 501f). The whole procedure is labeled "model-based", as opposed to the "design-based" approach (Spieß 2010: 120), in which observational probabilities are known, since the researcher assigned units of analysis with different selection probabilities.

Nonresponse weights can be combined with design weights in order to correct parameter estimation of an underlying target population, namely private households in Germany (in 2011 or 2012). Mean estimation for instance relies upon the estimator developed by Horvitz/Thompson (1952):

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{i=1}^{N} \frac{s_i}{\pi_i \cdot P(x_i h \in S)} x_i \tag{2}$$

[5] This notion is based on the assumption of a real population parameter, a mean for instance, which consists of the mean of participants $\bar{x}_r$ and non-participants $\bar{x}_n$. For instance, the nonresponse bias in a mean estimator $b_{\bar{x}}$ is a function of both the amount of variation between participants and non-participants as well as the share of nonrespondents (Bethlehem et al. 2011: 42):

$$b_{\bar{x}} = (\bar{x}_{response} - \bar{x}_{nonresponse}) \cdot \frac{n_{nonresponse}}{n_{total}} \tag{1}$$

In equation (2), π_i denotes the i-th individuals response probability and $P(x_i h \in S)$ denotes the sampling probability of the i-th person in strata h. s_i is a binary indicator taking on one for participation and zero otherwise (Kim/Kim 2007: 502).

This research note is structured as follows: Section 2 describes the sampling design of Sample J and Sample K and section 3 documents the prevalence of nonresponse. Section 4 reports the available characteristics of sampled addresses and section 5 reports regression models of non/response, which we use to generate appropriate nonresponse weights. Moreover, we document the balancing power of these weights and report some descriptive figures of the weights. Post-stratification (raking) of sample data – as one of the steps in the SOEP weighting procedure – is discussed in section 6. Section 7 reports characteristics of SOEP-"first wave weights", the result of a combination of design weighting, nonresponse adjustment and post-stratification provided for each of the different subsamples.

2 Sampling Design and Design Weights

The target population of the refreshment Sample J is the cross-section of private households residing in Germany in 2011 and the target population of Sample K are private households in 2012. Sample J was implemented in field from March to October 2011 and Sample K from March to October 2012. To ease fieldwork of face-to-face interviewing, we employed in both cases a clustered sampling strategy based on the ADM ("Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.") sampling frame that divides Germany into $\sim$ 53,000 spatial entities. Sample J uses a random sample of 307 and Sample K of 126 "sample points" that are both stratified for *Länder* (federal states), sub-*Länder* administrative regions, and a classification of municipalities according population size (SOEP 2012; SOEP 2013).

Within each sample point, random starting addresses were drawn for the following random walk procedure. In Sample J, interviewers collected 80 addresses out of which 30 were randomly chosen to be part of the sample[6]. In addition, an analysis of family names – the "onomastic procedure" (Humpert/Schneiderheinze 2013) – was performed by a specialized institute. Family names indicating a non-German origin in the 30-address-sample were then counted and the number of sampled "foreign" addresses was then increased by this number (SOEP 2012: 51). This is part of the a priori efforts to increase sampling of immigrants which are known to display low probabilities of participation. For each household in Sample J the design weight, as derived from sampling probabilities was calculated as follows:

$$w_d = \left(s_m \cdot \left(\frac{n_{m(p)}}{N_{m(p)}} \right)^{-1} + s_g \cdot \left(\frac{n_{g(p)}}{N_{g(p)}} \right)^{-1} \right) \cdot \frac{n_{m(p)} + n_{g(p)}}{N_{m(p)} + N_{g(p)}} \tag{3}$$

In equation (3), s_m (s_g) is a binary indicator denoting whether the household is coded on the basis of the given and the family name as having supposedly foreign (native) origin ($s = 1$ if yes, zero otherwise). The index p denotes the sample point an household belongs to. $n_{m(p)}$ ($n_{g(p)}$) is the number of migrant (native) households in the actual sample from samplepoint p, whereas $N_{m(p)}$ ($N_{g(p)}$ resp.) represents the number of migrant (native) households in the original address sample (result of every third address being recorded during random walk). The last element of the equation is a correction for the total number

[6]Note that these proportions varied a little during fieldwork, so that some sample points provided less than 80 addresses to choose from. Design weights were corrected for this fact.

of households being sampled in the sample point p, $n_{m(p)} + n_{g(p)}$, and the overall number of households recorded in the sample point, $N_{m(p)} + N_{g(p)}$.

In Sample K, again, 80 addresses were collected within each sample point and 36 were randomly chosen to be part of the sample. Contrary to Sample J, we did not assign different sampling probabilities to non/German households on the basis of an onomastic procedure. The sample is "self-weighting", i.e. every household in the target population has had the same chance to be sampled.[7] Hence, the design weight for households in Sample K is a constant factor.

3 Sample Size and Nonresponse

The actual computer-assisted personal interview (CAPI) of Samples J and K took place later and only after written announcement. Out of the 9,804 households in the gross sample of Sample J, 32% (3,136) were interviewed partially or completely during the sampling period. Within the non-participating households, 319 households were classified as "quality neutral non-response"[8], and are not analyzed any further. The overall response rate within this reduced gross sample amounts to 33% (AAPOR Non-Response Definition RR2, see (AAPOR 2011)).

In Sample K, a total of 4,536 households were sampled to participate in the survey and 1,526 of these households were successfully interviewed. Within the non-participating households 139 households were classified as "quality neutral" non-response. The overall response rate within Sample K amounts to 35%. Figure 1 displays response rates according to the *Länder*- and the county-level.[9] As can be seen, nonresponse displays cross-sectional variation. Explaining this variation will be one main task of this paper.

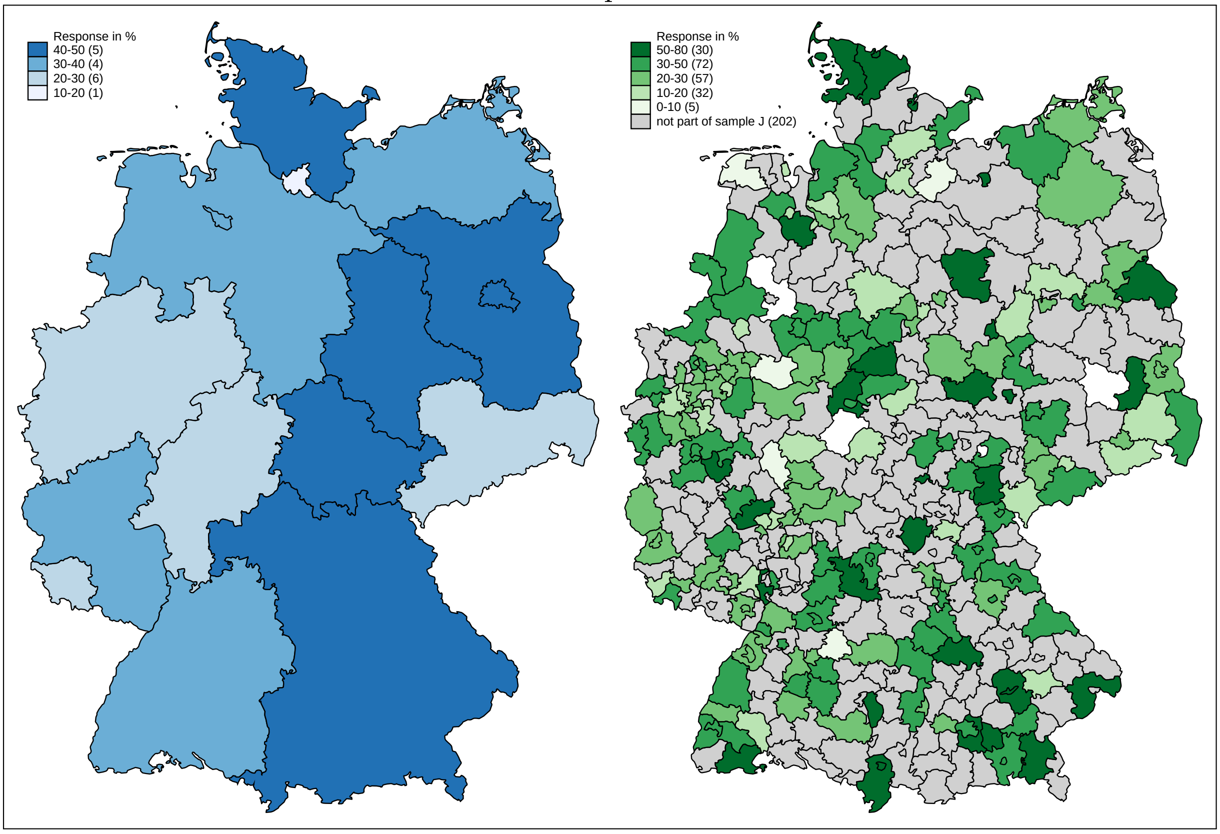
[7] Sampling designs holding this characteristic are also referred to as "EPSEM": Equal Probability Selection Method.

[8] This means for instance, that false addresses were recorded, persons decreased, moved abroad, or interviewers were unable to complete sampling in time, due to illness, for instance.

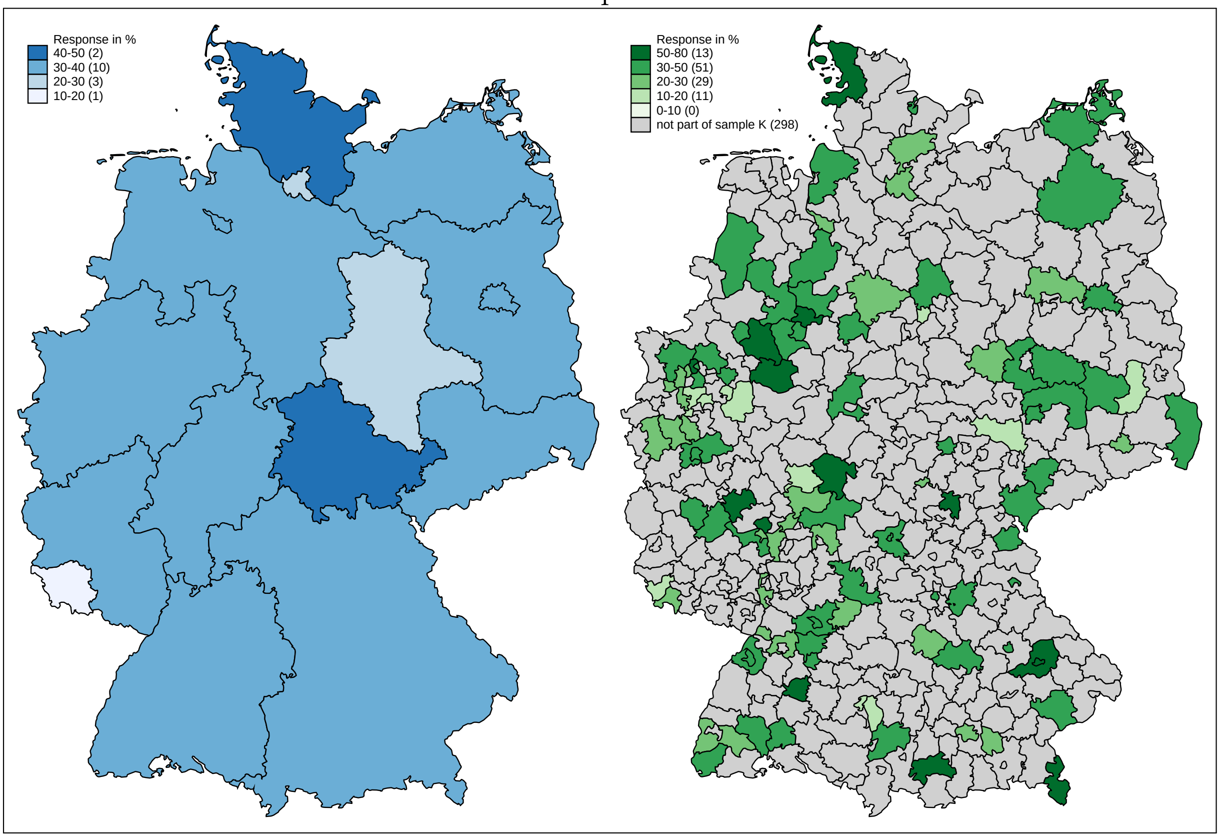
[9] As displayed by the map on the right, a large number of German counties were not part of the refreshment Samples J and K (gray areas).

Figure 1: Response Rates in Samples J and K by *Länder* and Counties

Sample J



Sample K



Note: AAPOR Response Rates 2 (RR2) (AAPOR 2011).

4 Correlates of Nonresponse: Data Sources

A model-based estimation of response-propensities that lend themselves as the basis for weighting variables compensating for selective participation rates requires observable information on both responding and non-responding households. This paper makes use of information from different sources to model nonresponse. Due to spatial constraints, this section gives only a very brief overview over the different sources and the variables, Appendix A provides detailed information about the expected effects of variables on response probabilities. Note, that the focus in analysis lies upon the consistent estimation of response propensities, not on the theoretical interpretation of effects (Spieß 2010: 123). Furthermore, a distinction has to be made between variables available for individual households and spatial data linked to the households on the basis of regional identifiers. Implying causal effect at the individual level because of significant relationships at the aggregate level raises the problem of ecological fallacies (McGaw/Watson 1976: 134f). Therefore, caution is needed in interpretation. At the end of the section we provide a table summarizing all variables in their original form.

4.1 Addresses: Field Work Information

During address sampling, interviewers collected information about households and their environment, such as the supposed migration background of households as indicated by family names. Immigrant households should on average perceive higher costs of participation because of difficulties concerning language and therefore participate less frequently (e.g. Bethlehem et al. 2011: 64). Although questionnaires are available in several languages for the SOEP, there is no guarantee this fact is known to households when a decision for or against participation is made. Furthermore, the type of house (e.g. flat in multi-story building vs. individual house) was recorded. This variable contains useful information about the living standard of sampled households. Residents in more expensive individual houses (possibly house owners) are supposed to be more prone to participation than people living in flats (Durrant/Steele 2009: 376), as for wealthier individuals have been reported to show higher participation probabilities (see. Abraham et al. 2006: 693f). Another variable used is the type of neighborhood. Households in "accommodation only" districts are expected to be more readily participating than those in more isolated industrial/commercial areas (Durrant/Steele 2009: 375). Also, the size of the community (number of inhabitants) was coded[10].

4.2 Neighborhood: Microm Data

The next data source used is a dataset provided by the private enterprise "microm GmbH" and may be used by guests and staff of the SOEP (Goebel et al. 2007). It contains detailed local and regional information about the social structure and environment/neighborhoods of households in Germany. Variables are available at different levels of aggregation, ranging from the household-cell-level (few households grouped together), over market-cells (ca. 470 households per cell) to 8-digit postal code districts (ca. 500 households per district). Microm-Data therefore provides very fine-grained regional data for analysis. The variables used here mainly measure the social structures of households (e.g. age, family structure, education, migration) as well as the economic situation of households (e.g. unemployment, purchasing power).

4.3 Municipality: Regional Information from the Federal Statistical Office

As a joint project of the Federal Statistics Office with its subnational counterparts, the "Regionaldatenbank Deutschland" (regional database Germany) provides register data on different levels of aggregation. For analysis of nonresponse in this paper, variables compiled at county level as well as at municipality level were obtained. The variables divide into three topics: data from the 2009 general election (turnout, share for different parties), age structure and distribution of different dwelling forms.[11].

4.4 County: INKAR Database

The database "Indikatoren und Karten zur Raum- und Stadtentwicklung in Deutschland und in Europa" (INKAR) is provided by the Federal Institute for Research on Building, Urban Affairs, and Spatial Development and contains official register information on economic issues (e.g. prices for building grounds, household income, welfare benefits) as well as the nature of inhabitants (e.g. educational data) of regional entities in Germany.

[10] Steps were as following: $< 2k$; $2k - 5k$; $5k - 20k$; $20k - 50k$ $50k - 100k$; $100k - 500k$; $> 500k$

[11] For further information, see the link under Regionaldatenbank Deutschland (2012) in the bibliography.

Variables were available at the county-level and for NUTS 2[12] regions and compiled in 2009[13].

4.5 Additional Data Sources

Two more variables were obtained from the comparative research project “Deutscher Lernatlas” on conditions of learning quality at the regional level. Data can be downloaded freely without registration[14]. Variables related to integration of citizens in societal activities (amount of volunteering) and political activity (partisanship) available at the county level were extracted for this paper.

[12]NUTS 2 is a statistical region used in cross-country comparison by European Union Statisticians.

[13]For additional information on variables and technical issues, see INKAR (2011) in the bibliography.

[14]See Lernatlas (2011) and `http://www.deutscher-lernatlas.de/de/ergebnisse/daten.html` [visited the 04th June 2014].

Table 1: List of Variables used in Analysis of Nonresponse of Samples J and K

Variable	Source	Type	Values/ Range	level	Year	expected effect
migrant (family name)	field information	binary	0= no 1= yes	household	2011/2012	negative
type of house	field information	ordinal (4 steps)	1= individual 4= high multi-story	household	2011/2012	negative
business intensity (district)	field information	ordinal (5 steps)	1=residential district 5=industrial zone	household	2011/2012	negative
municipality size	field information	ordinal (6 steps)	1= <2k inh. 7= > 500k inh.	municipality	2011/2012	negative
business intensity (street)	Microm	ordinal (6 steps)	1 = accommodation only 6 = business only	street level	2011/2012	negative
mean age of heads of houses	Microm	ordinal (8 steps)	1 = <35 8 = 65+	house cells	2011/2012	negative
household structure	Microm	ordinal (9 steps)	1= mainly single persons 9= mainly families with children	house cells	2011/2012	positive
children per household	Microm	ordinal (9 steps)	1= lowest value 9= highest value 6= average	house cells	2011/2012	positive
status (socio-economic)	Microm	ordinal (9 steps)	1= loewest status 9= highest status 5= average	house cells	2011/2012	positive
share of college graduates	Microm	ordinal (7 steps)	1= below 2% 7 = above 35%	street level	2011/2012	positive
exclusive housing environment	Microm	binary	1=yes 0=no	house cells	2011/2012	negative
purchasing power	Microm	metric	100= national average	market cells	2011/2012	positive
share of Turkish immigrants	Microm	metric	-	market cells	2011/2012	-
share of eastern European immigrants	Microm	metric	-	market cells	2011/2012	-
turnover in accommodation (mobility)	Microm	ordinal (9 steps)	1= lowest value 9= highest value 5= average	market cells	2011/2012	negative
balance of accomod. turnover (mobility)	Microm	ordinal (9 steps)	1= extr. negative 9= extr. positive 5= balanced	market cells	2011/2012	positive
unemployment	Microm	ordinal (7 steps)	1= lowest 7= highest 4= national average	8-digit postal codes	2011/2012	negative
prices for building grounds	Inkar	metric	in €/m^2	county	2009	positive
average household income (per person)	Inkar	metric	in €	county	2009	positive
GDP/capita	Inkar	metric	in 1000 €‘s	county	2009	positive
welfare benefits for renting expenses	Inkar	metric	in €	county	2009	positive
med. doctors per 100k inhabitants ratio	Inkar	metric	-	county	2009	positive
share of high school graduates	Inkar	metric	-	NUTS 2	2009	positive
share of college graduates	Inkar	metric	-	NUTS 2	2009	positive
electoral turnout in 2009 general election	Statistics Office	metric	-	municipality	2009	positive
vote share for SPD	Statistics Office	metric	-	municipality	2009	negative
vote share for CDU/CSU	Statistics Office	metric	-	municipality	2009	negative
vote share for FDP	Statistics Office	metric	-	municipality	2009	negative
vote share for *Alliance '90/ The Greens*	Statistics Office	metric	-	municipality	2009	negative
vote share for *The Left*	Statistics Office	metric	-	municipality	2009	negative
vote share for small parties	Statistics Office	metric	-	municipality	2009	positive
share of small flats (1-2 rooms)	Statistical Office	metric	-	municipality	2010/2011	negative
share of big flats (6+ rooms)	Statistical Office	metric	-	municipality	2010/2011	negative
share of 18-25 aged	Statistical Office	metric	-	municipality	2010/2011	-
share of 25-35 aged	Statistical Office	metric	-	municipality	2010/2011	-
share of 35-45 aged	Statistical Office	metric	-	municipality	2010/2011	-
share of 45-55 aged	Statistical Office	metric	-	municipality	2010/2011	-
share of 55-65 aged	Statistical Office	metric	-	municipality	2010/2011	-
share of elderly (65+)	Statistical Office	metric	-	municipality	2010/2011	-
share of people active in non-profit org.	Lernatlas	metric	-	county	2008	positive
quota of party members	Lernatlas	metric	-	county	2009	positive

4.6 Multiple Imputation and Data Coding

Some of the variables obtained contained missings. In the majority of cases, all values for all variables for one source were missing for a spatial unit (county, municipality). However, none of the households yield complete missings. In other words, missings do not cluster for one particular set of households. Furthermore, overall missingness was low, as can be seen from tables 2 reporting the prevalence of missing data by groups of indicators and *Länder*.

Table 2: Share of Missings in Sample J and Sample K for Different Variables by *Länder*

Variables *Länder*	Field inf. (Address)	Microm (House Cells)	Microm (Others)	INKAR (County)	Statistical Office (Municipality)	Lernatlas (County)
	J/K	J/K	J/K	J/K	J/K	J/K
Schleswig-Hol.	.0383/.0111	.0192/0.667	.0000/.0111	.0000/.0000	.0000/.0000	.0000/.0000
Hamburg	.0039/.0347	.0157/.0000	.0000/.0000	.0000/.0000	.0000/.0000	.0000/.0000
Lower Saxony	.0851/.0394	.0224/.0324	.0045/.0000	.0000/.0000	.0000/.0000	.0000/.0000
Bremen	.0833/.0278	.0139/.0000	.0000/.0000	.0000/.0000	.0000/.0000	.0000/.0000
Northrhine Westph.	.0567/.0283	.0198/.0062	.0042/.0000	.0000/.0000	.0000/.0000	.0000/.0000
Hesse	.0537/.0864	.0358/.0093	.0000/.0000	.0000/.0000	.0000/.0000	.0000/.0000
Rheinland-Palatinate	.0532/.0417	.0298/.0370	.0043/.0046	.0681/.0000	.0000/.0000	.0000/.0000
Baden-WuerT.	.1597/.0574	.0269/.0481	.0040/.0046	.0000/.0000	.0000/.0000	.0269/.0667
Bavaria	.0491/.0351	.0267/.0263	.0014/.0093	.0000/.0000	.0000/.0000	.0000/.0000
Saarland	.1544/.0833	.0074/.0000	.0000/.0000	.0000/.0000	.0000/.0000	.0000/.0000
Berlin	.0663/.0000	.0141/.0243	.0020/.0174	.0000/.0000	.0000/.0000	.0000/.0000
Brandenburg	.0132/.0000	.1026/.0741	.0596/.0000	.0000/.0000	.0000/.0000	.0000/.0000
Mecklenb.-Vorp.	.0238/.0139	.0000/.0000	.0000/.0000	.0000/.0000	.0000/.0000	.0000/.0000
Saxony	.0233/.0159	.0388/.0238	.0078/.0079	.0000/.0000	.0000/.0000	.0000/.0000
Saxony-Anhalt	.0257/.0159	.0478/.0069	.0000/.0000	.0000/.0000	.1875/.0000	.0000/.0000
Thuringia	.0148/.2407	.0221/.1296	.0000/.0000	.0000/.0000	.0000/.0000	.0000/.0000
Total	.0658/.0039	.0276/.0027	.0052/.0004	.0039/.0000	.0058 /.0000	.0041 /.0079

Note: If variables were grouped together (e.g. INKAR) in columns, the variable with the highest share of missings was used for calculation. Thus, some variables in those groups are less incomplete. Variables not mentioned here were already complete. Values were calculated using all observations.

For analysis of nonresponse and the generation of weights, it is necessary to have complete observations. Otherwise, observations will be omitted from regression and weights cannot be estimated for those observations. Therefore, missing observations were imputed using "multiple imputation by chained equations" (Royston 2009). To account for imputation uncertainty, ten different predictions were made (White et al. 2011: 378). Furthermore, the whole procedure was implemented ten times with different starting values (Horton/Lipsitz 2001: 248). As a result, ten different complete datasets are available for analysis taking the uncertainty of multiple imputation into account via appropriate statistical routines (White et al. 2011: 377).

After imputation, variables were transformed for analysis. Continuous Variables were categorized accounting for special features of their distribution (e.g. multiple modi, outliers). In the majority of cases, this led to three distinct categories for each variable. In general, the middle category served as a reference group in regression. Some continuous variables, however, were dichotomized during transformation. Ordinal indicators with several categories (e.g. socio-economic status) were recoded to two or three categories in order to produce more qualitatively distinct groups. Using categorized variables and their respective binary indicators in regression has several advantages in this context. First of all, non-linear effects are controlled for, because for each group individual parameters are estimated. Furthermore, the categorization prevents the estimation of extreme probabilities very close to zero or one because of single outliers on a variable. This is necessary in order not to inflate the estimated weights inappropriately (Spieß 2010: 122; Valliant/Dever 2011: 116). Finally, interpretation and comparison of coefficients is more convenient this way (Zaslavsky et al. 2002: 487).

5 Modeling Nonresponse and Nonresponse Weights

To model the households' nonresponse propensities in Samples J and K, logit regression was performed for different combinations of covariates using statistical routines to account for imputation uncertainty. Furthermore, we used robust estimation of standard errors in order to account for the possibility of heteroscedasticity and non-independent observations in sample points (see White 1980, Spieß 2010). The identifier of sample point membership of households was used as the cluster variable. Not doing so would yield the risk of estimating too large or too small standard errors, the latter being even more threatening to valid inferences. In addition to the variables mentioned above, dummy-variables for the *Länder* were included.

5.1 Nonresponse Model Sample J

All in all, 9,479 households were used in every model[15]. Figure 2 displays coefficients and their 95% confidence interval calculated using the standard errors.

The full model uses all variables available as covariates. The second, reduced model was estimated using only those variables that exert a significant effect ($\alpha = 5\%$-level). Both models show the relative independence of response propensity to characteristics of the neighbourhood. Only a small number of the predictors reaches statistical significance in the models. Thus, the reduced model is a lot more parsimonious.

Table 3: Fit Values for Estimated Models of Sample J

	Full model	Reduced model
pseudo-R^2	.05	.04
error rate	.32	.33

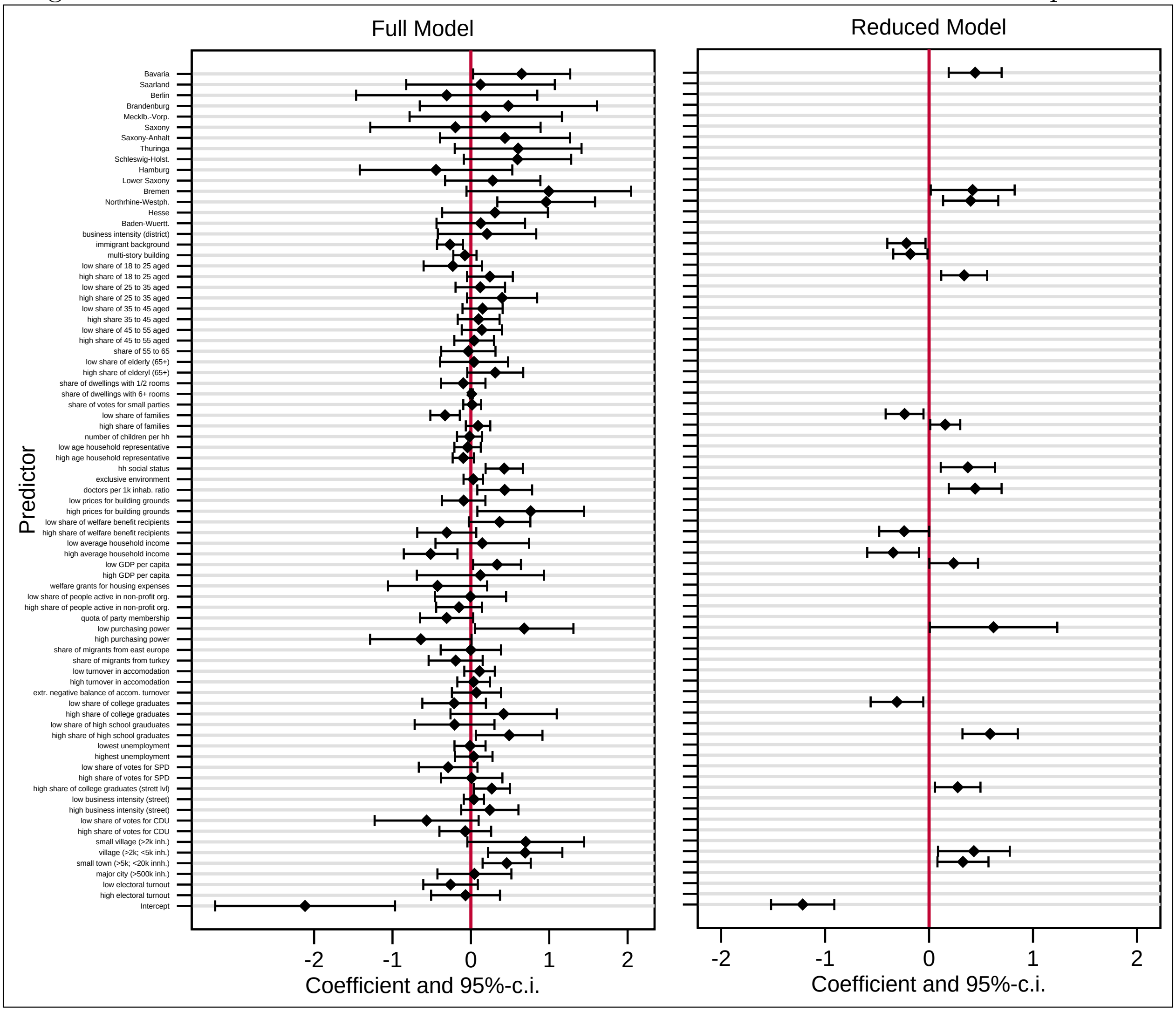
Note: For calculation of the error rate, see Gelman/Hill (2007: 99).

Regarding the different criteria for model fit of both models as reported in Table 3, no substantial differences arise[16]. Although the full model fares slightly better in comparison

[15] From the 9,804 sampled households, 6 asked to delete their data and 319 of the non-responding households were classified as "quality neutral" and are not analyzed here.

[16] Note that due to the lack of independence between observations in multiply imputed datasets, likelihood-based measures of model fit cannot be calculated appropriately. Therefore, the values for the reported *Mac-Fadden's Pseudo-R^2* were calculated using normal logistic regression with cluster-robust standard errors. Due to the small fraction of missing data, however, estimated parameters differed only marginally (from the third decimal place on).

Figure 2: Coefficients and Confidence Intervals for the Estimated Models of Sample J



Note: The dependent variable was coded 1 for participation and 0 for non-participation. Number of observations in both models n = 9,479.

to the null model, both models do not have very good fit values. Nevertheless, the results are comparable to other works modeling nonresponse.

Keeping in mind the quality of sampling these results can be understood in a positive manner. A wide array of different variables have been tested for their influence on response probabilities and only few of them reach significance. Therefore, participation across groups indeed seems to be governed a lot by chance and in many aspects, respondents and nonrespondents possibly do not differ very much.

Turning to the individual coefficients, differences between the two models remain small. Many of the significant coefficients in the reduced model are also significant when estimated in the full model[17]. Moreover, in the majority of cases estimated coefficients yield the

[17]Note: Significant variables from the full model not included in the reduced model were very sensitive

expected sign, although most of them do not reach statistical significance in the full model, sometimes only by a small margin. Therefore the reduced model seems to be more suited for analyzing nonresponse.

As expected, the educational level of a region's inhabitants relate to response probabilities. As can be seen by the three coefficients for college graduates (2 variables) and high school graduates, response probabilities seem to be higher in areas with higher average education. The covariates capturing the social structure of inhabitants also fare quite well. Both variables for household structure indicate that households in areas populated by families display high probabilities of participation, as opposed to households in areas dominated by single households which yield lower probabilities. Furthermore, households with members of supposedly foreign origin and households situated in flats in multi-story-buildings are more reluctant to participate in the survey than German households and residents of individual houses, possibly located in wealthier suburbs. A high medical doctors per inhabitants ratio as an indicator for "advantaged areas"(Johnson et al. 2006: 707f) also yields the expected positive effect. Finally, participation probabilities are significantly higher in smaller cities, since the two of the three corresponding coefficients show the expected signs.

The picture for variables relating to the economic conditions of households in an area, however, is mixed. Only a high share of people entitled to welfare benefits seems to relate to participation probabilities in the expected negative way. The other coefficients are estimated opposite from what was expected, thereby indicating higher response probabilities for households in economical weaker areas. Theory would classify those regions as exhibiting "concentrated disadvantage" and they were expected to reduce participation probabilities because of inhabitants of such areas being less integrated into civic society (ibid.: 707f). Among them are some of the biggest coefficients (e.g. "low purchasing power"). The explanation of these effects remains unclear, especially since any causal explanation in this research settings may fall prey to ecological fallacies. However, Durrant/Steele (2009) point out that overall findings on the effect of such variables have been mixed in the past. Nevertheless, it becomes evident that controlling for economical well-being of households with multiple variables is reasonable.

Significant effects are also observed among the controls included without any specific expectations. Of the binary indicators for the *Länder*, three yield significant positive effects.

to different model specifications and therefore excluded.

Also, a high share of inhabitants aged from 18 to 25 years in a municipality positively relates to participation probabilities. An ad-hoc explanation would point to the fact, that people aged from 18 to 25 supposedly have not had the chance to participate in a lot of surveys because of their young age. Therefore saturation effects might not be as strongly developed in areas with a high share of young adults.

5.2 Nonresponse Model Sample K

In Sample K, a total of 4,397 households were used in every model[18]. Figure 3 displays coefficients and their 95% confidence interval calculated using respective standard errors.

Again, only a relatively small fraction of the predictors reaches statistical significance in the full model, signaling the relative balance of the realized net sample compared to the gross sample. The reduced model is a lot more parsimonious. Regarding the two criteria for model fit reported in table 4, only small differences between the full and the reduced model arise. According to the pseudo-R^2, the full model fares slightly better in comparison to the reduced model. However, there is almost no difference between error rates.

Table 4: Fit Values for Estimated Models of Sample K

	Full model	Reduced model
pseudo-R^2	.04	.03
error rate	.33	.34

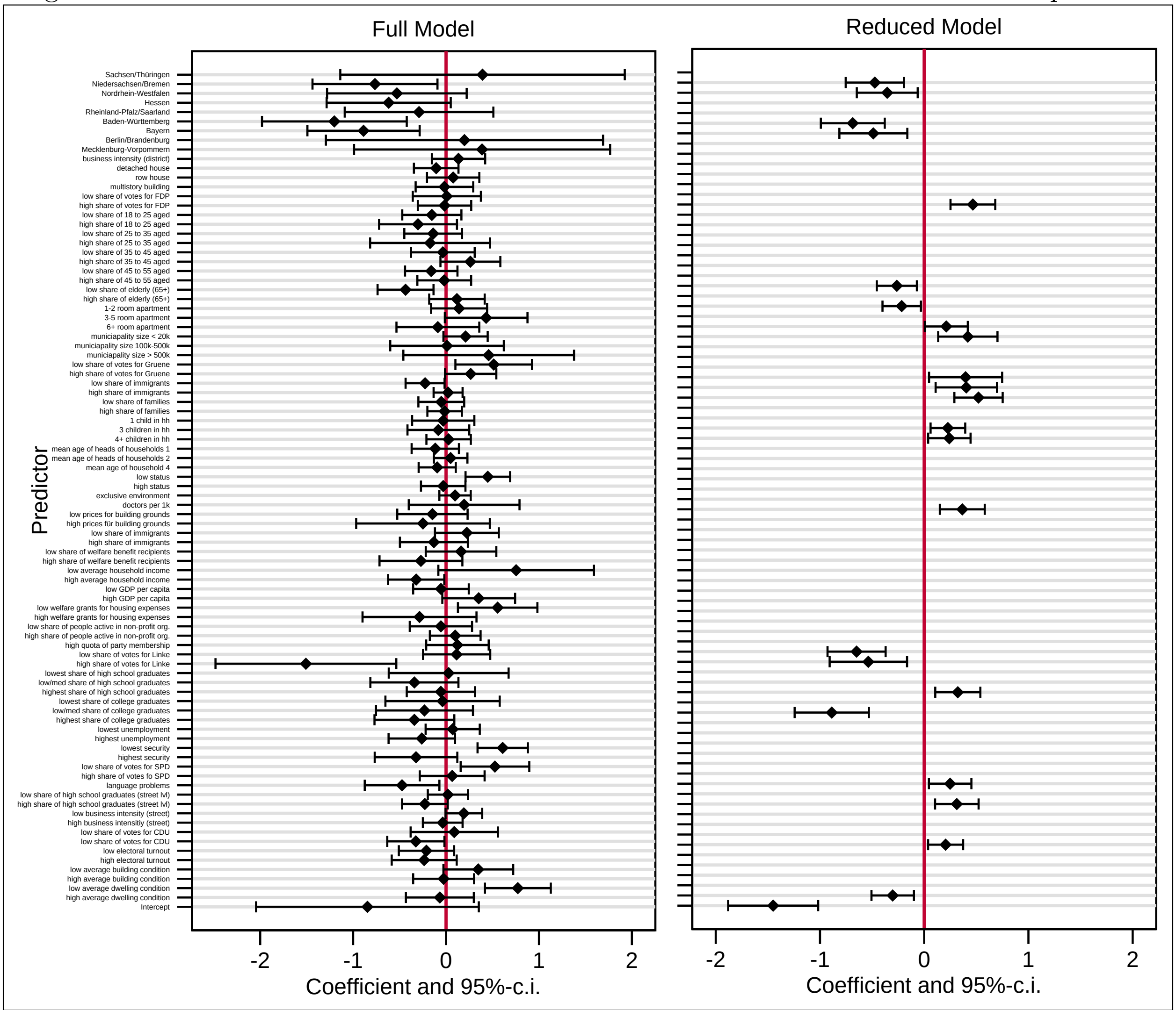
Note: For calculation of the error rate, see Gelman/Hill 2007: 99.

Significant effects are found with regard to one of the coefficients capturing the social integration of citizens. As expected, participation probabilities are significantly higher in areas with high rates of party membership.

A high share of election votes for the party *Die Linke* (*The Left*) in an area yields an negative effect on participation probabilities. By contrast, high shares of votes for the FDP are related to higher participation probabilities. Coefficients for *The Greens* suggest ambivalent relations as both areas with high and low shares of votes (compared to a middle category) show significant positive effects on participation probabilities. Finally, and in line with our expectations, areas with low shares of votes for the socio-democratic party

[18]The gross sample covered 4,536 households from which 1 household moved abroad, 6 households were deceased, and 132 of the non-responding households were classified as “quality neutral”. These 139 households are not analyzed any further.

Figure 3: Coefficients and Confidence Intervals for the Estimated Models of Sample K



Note: The dependent variable was coded 1 for participation and 0 for non-participation. Number of observations in both models n = 4397.

(SPD) display higher participation rates. The coefficient for overall election turnout in 2009 is estimated contrary to what was expected, indicating lower participation probabilities for areas with a high election turnout.

The picture for variables relating to the socio-economic conditions of households in an area, however, is inconclusive. As expected, households in areas with low rates of unemployment display significantly higher response rates. Yet, contradictory results are obtained for other variables capturing the inhabitants' socio-economic conditions. Participation probabilities seem to be increased in areas with a high share of low status households. Furthermore, response probabilities are significantly higher in areas with a low share of households entitled to welfare grants for housing expenses.

Contrary to our findings on Sample J, coefficients capturing the educational level of a region's inhabitants do not affect response probabilities significantly in Sample K. As can be seen by the coefficients for college and high school graduates, response probabilities do not seem to be related to the educational level of an area's inhabitants.

Finally, a low share of inhabitants in municipalities aged 45-55 and 65+ years positively relates to participation probabilities. These results are in line with other studies (e.g. Keeter et al. 2006). Furthermore, there are significant differences in response propensities between federal states. Lower Saxony/Bremen, Northrhine-Westphalia, Baden-Württemberg and Bavaria yield significant negative effects, i.e. lower response probabilities compared to other *Länder*. Since all of these states are located in West-Germany, the results may portend to a more general difference in response propensities between East- and West-Germany.

5.2.1 Generation of Nonresponse Weights

The reduced models from section 5 were used to predict participation probabilities in Samples J and K. The models are suitable for analysis since their effects have proven to be robust among different regression specifications and they do not contain insignificant predictors artificially increasing the variation of estimated probabilities and thereby also nonresponse weights[19]. Table 5 displays a comparison between actual response rates and mean estimated response probability by sample point in Samples J and K. While the model fit values did not indicate very strong differences between regional characteristics and individual response behavior, the correlations between the two variables shows that prediction of response rates at the sample point level is reasonably good, especially since model specification was intended to reduce the estimated probabilities variation. Thus, the categorization of variables clearly succeeded in limiting the estimated response probabilities, since actual variation in response rates was substantially larger than estimated. The nonresponse weights for further analysis are calculated as inverse response probabilities.

Table 5: Comparison of Estimated and Actual Response Rates by Sample Point

Response Rates	Minimum	25%-Quantil	50%-Quantil	75%-Quantil	Maximum
Sample J					
observed (x_{obs})	.00	.19	.30	.46	.88
estimated (x_{est})	.14	.27	.32	.40	.61
Sample K					
observed (x_{obs})	.00	.25	.33	.42	.69
estimated (x_{est})	.19	.29	.35	.40	.61

Note: At the sample point level the two variables correlate with $\rho_{(obs;est)} = 0.53$ in Sample J and with $\rho_{(obs;est)} = 0.68$ in Sample K.

Table 6 contains a description of the raw estimated weights. Alongside with mean weights and standard deviations, quantiles are also included and weights are tabulated according to specific values of selected covariates.

The table shows how some groups receive higher weights, while others are weighted down. Note, however, that for some variables differing a lot in mean weights and most of the weights located close to the mean, minimum and maximum values do not differ a lot (e.g. migrants in Sample J). This may be regarded as a consequence of many different variables

[19] The range of estimated probabilities for the full model in Sample J, for instance, was about 30 percentage points larger compared to the reduced model. Estimation of probabilities in the full model without any categorization of variables yields a 34% larger range of weights compared to the reduced model.

Table 6: Raw Estimated Nonresponse Weights by Different Variables

	Min.	Quantiles 10%	25%	50%	75%	90%	Max.	Mean	SD
Total Sample J	1.48	2.10	2.48	3.11	3.91	5.03	10.03	3.36	1.19
immigr. background = yes	1.54	2.54	3.14	4.09	5.07	6.18	10.03	4.27	1.47
immigr. background = no	1.48	2.06	2.45	3.05	3.75	4.73	9.90	3.23	1.08
high share of families	1.48	1.99	2.22	2.65	3.21	3.71	7.12	2.77	0.71
low share of families	1.63	2.33	2.84	3.79	4.79	6.08	10.03	3.95	1.39
highest status group	1.52	2.14	2.50	3.14	3.95	5.06	10.03	3.40	1.19
lowest status group	1.48	1.88	2.21	2.79	3.68	4.54	8.64	3.08	1.19
Total Sample K	1.54	2.15	2.47	3.01	3.50	4.14	6.84	3.06	0.77
high share of immigrants	1.54	2.26	2.58	3.08	3.64	4.17	6.84	3.17	0.79
low share of immigrants	1.54	2.06	2.36	2.89	3.43	4.02	5.73	2.97	0.77
high share of families	1.54	2.06	2.35	2.68	3.26	3.98	5.07	2.86	0.69
low share of families	1.82	2.44	2.83	3.34	3.86	4.60	6.84	3.41	0.79
highest status group	1.78	2.23	2.63	3.22	3.80	4.18	5.73	3.24	0.77
lowest status group	1.54	1.95	2.32	2.71	3.20	3.58	5.05	2.79	0.69

Note: All values were rounded to the second decimal place. Values for all observations (all ten imputed datasets) displayed.

being taken into account and those variables not being multicollinear.

The estimated weights even after having been reduced in variation through regression design still cover a very wide range. Especially the difference between the values of the 90%-quantile and the corresponding maximum in the rows of table 6 is rather extreme, especially in Sample J. Therefore, some outliers must be present in the data. As mentioned before, estimating extreme weights can be harmful. In effect, variation of estimates may increase substantially (van Goor/Stuiver 1998: 295). Therefore, additional measures need to be taken. Trimming weights to be more equally distributed as has been done elsewhere (e.g. Peytchev et al. 2011: 149f) seems reasonable. The trimming of weights, however, may result in a small loss of efficiency. For trimming the weights, decisions have to be taken regarding the desired range of weights[20]. This decision is best based on the ratio of estimated weights. In Sample J, it was decided for the ratio of weights in relation to the minimum weight not to exceed 4.5. This basically affected all households having received a weight above approx. 6.65. As can be seen by the raw estimated weights in Table 6, only a small fraction of observations was affected by this. Transformation used a logarithmic function, as for it has a decreasing slope, thereby exerting stronger transformation for more extreme outliers than for outliers close to the threshold. In Sample K, we aimed for household weights not to exceed 1.75 times the weight's mean:

[20]Theoretical guidance on appropriate figures is sparse, for an example for the difficulties of trimming, see van Goor/Stuiver (1998).

$$w_{trimmed} \begin{cases} 1.75 * \bar{x}_w + 0.5 * ln(w - 1.75 * \bar{x}_w + 1), & \text{if } w > 1.75 * \bar{x}_w \\ w, & \text{otherwise} \end{cases}$$

This affected only twelve households having received a weight above approx. 5.2. Hence, trimming succeeded as variation in weights was decreased while only few weights have been corrected and therefore holding loss of efficiency at a minimum level.

Table 7: Comparison of Weighted and Unweighted Estimates and Reduction of Bias in Sample J and Sample K

Variable		Real value	Estimated Bias in Percent:[a] Unweighted		Weighted		Reduction of Bias
GDP per cap. (in 1000)							
	J:	29.82	3.00	***	0.99		67%
	K:	29.73	3.11	***	0.21		93%
share of people entitled to welfare benefits (per 1k inh.)							
	J:	85.27	3.07	***	0.41		86%
	K:	86.25	3.23	**	0.01		98%
household income							
	J:	1574.40	1.40	***	0.04		97%
	K:	1578.05	0.90	**	0.27		70%
share of high school graduates							
	J:	58.95	1.06	***	0.04		96%
	K:	58.80	0.60	**	0.00		93%
share of college graduates							
	J:	26.49	0.32		0.20		39%
	K:	26.59	0.43		0.19		55%
med. doctors per 100k inhabitants ratio							
	J:	171.51	1.47	**	0.04		97%
	K:	171.50	2.44	**	0.56		77%
prices for building grounds per m^2							
	J:	184.81	7.55	***	4.54	**	40%
	K:	183.86	8.15	***	0.00		99%

[a] Note: *'s indicate result of two-tailed t-test for differences in means (difference from real value). $* \rightarrow p < .1$; $** \rightarrow p < .05$; $* * * \rightarrow p < .01$;

After having estimated nonresponse weights, testing them is feasible using the known data. In order to test whether weighting may reduce bias in estimation, unweighted and weighted mean estimates using only participating households are compared to the actual value estimated using data for all households of the gross sample. The product of nonresponse and design weights was calculated to constitute the combined weight used in this analysis (and the design weights in the gross-sample only). Table 7 shows the comparison mentioned for several of the variables used in analysis[21].

Bias was calculated as percent deviation from the actual value[22]. Reduction in bias was calculated as one minus the ratio of the weighted and the unweighted estimate‘s bias and is reported without signs.

As can be seen, the weighted estimates of the realized sample for all variables are closer to the actual values observed in the gross sample. Thus, estimation was improved using the estimated nonresponse weights. The results of the t-test provide further evidence for the success of weighting. Almost all estimated unweighted means differ significantly from the actual value of the full sample. However, after adjustment for nonresponse, the bias in estimates is not significant for all but one variable, which is significant at the $\alpha = 5\%$-level. To put it another way, this is what would be expected in five percent of the cases even if the respondents were a perfect random sample of the combined set of nonrespondents and respondents. Moreover, bias is nevertheless reduced substantially for this variable, as reported in out right column of table 7. All in all, weights have helped to decrease relative bias for all variables reported here and thereby improved estimation.[23]

[21] Note that this analysis has been performed with the first of the imputed datasets. Household-specific means of estimated weights were calculated and used in analysis. Finally, trimming was implemented only after this procedure.

[22] The actual value refers to the estimated mean using all households and the design weights.

[23] The impact of weights can also be assessed using other methods of inference. Drawing upon the test proposed by DuMouchel/Duncan (1983), weights can be tested for their significant impact on regression parameters. It was originally developed for the assessment of model specification and tests parameters in weighted and unweighted regression against each other. The result of this test is designed to be a guidance for researchers on when incorporating weights into analysis is necessary and when not (ibid.: 593). The test itself uses weights and interactions with the predictor. The resulting coefficients are then assessed in a F-test; significance indicates that weights cannot be ignored since the effect of interactions and weights are supposed to be zero when weights are to be ignored.The test is available using the `wgttest` module in Stata developed by Ben Jann, University of Bern.

6 Post-Stratification

In addition to the reported design and nonresponse weighting adjustment, we corrected household weights of both Sample J and Sample K to meet known marginals from the underlying target populations. In standard post-stratification or cell weighting procedures, weights are adjusted so that given sample totals fit to known population totals in various strata (cells) (Kalton/Flores-Cervantes 2003). SOEP uses *raking* (also referred to as "iterative proportional fitting", (Deming/Stephan 1940)) for fitting a number of marginal distributions (Lohr 2010: 344).

Total marginal values derived from the German Microcensus were used in the raking procedure in which weights were corrected to meet benchmarks of the underlying target population in 2011 (Sample J) and 2012 (Sample K). The Microcensus is conducted by the Federal Statistical Office of Germany (FSO) and is a one-percent sample of the German resident population. The survey seems to be ideally suited as a benchmark because of its large sample size[24] and an obligation to provide information for most of the questions. Hence, comparatively high-quality population estimators are expectable.

Table 8: Population Characteristics Used in the Raking Procedure

Variable	Values	level	Source
household size	1 / 2 / 3 / 4 / 5+	household	FSO
Länder	Baden-Württemberg / Bavaria / Berlin and Brandenburg / Hesse / Lower Saxony and Bremen / Mecklenburg-Vorpommern / North Rhine-Westphalia / Rhineland-Palatinate and Saarland / Saxony / Saxony-Anhalt / Schleswig-Holstein and Hamburg / Thuringia	household	FSO
municipality size	< 20.000 / 20.000-100.000 / 100.000-500.000 / > 500.000	household	FSO
house ownership	yes / no	household	FSO
number of employed household members[a]	0 / 1 / 2 / 3 / 4+	household	FSO
recipient of unemployed benefit II (ALG II)[b]	yes / no	household	FSO
household type[c]	1 Adult / 2 Adults / 2 Adults, 1 child / 2 Adults, 2 children / etc.	individual	SOEP
gender	male / female	individual	FSO
citizenship	German / Foreign	individual	FSO
age	0-15 / 15-20 / 20-25 / 25-30 / 35-40 / 40-45 / 50-55 / 55-60 / 60-65 / 65+	individual	FSO

[a] Separated by East and West Germany for Sample J.

[b] Separated by East and West Germany.

[c] At individual level, the post-stratification of household type is also used to balance differences between the total number of household members and the number of individuals actually interviewed. Those differences result from partial unit-nonresponse in households.

[24] For example, the Microcensus of 2012 provides information on 337.600 households with 688.900 individuals (Statistisches Bundesamt 2013).

Six variables and their respective marginal distributions were used to rake the combined design and nonresponse weights of samples J and K at household level (see table 6). The first four represent "standard" characteristics of SOEP ranking procedures. Since preliminary analysis showed differences between Sample J and K on the one hand and official statistics on the other hand, the last two characteristics, namely the number of employed persons per household and the receipt of unemployed benefit II (ALG II), were added to that list.

The post-stratification procedure completes the three step process of calculating "combined first wave weights" for new SOEP samples. These first wave weights are therefore the result of a combination of design weighting, nonresponse adjustment, and post-stratification. First wave weights are available for all SOEP subsamples A to K. They are of special importance as they serve as base weights for the calculation of both longitudinal weights and cross-sectional weights from wave 2 onwards. The cross-sectional household weights are stored in the variables `BBHHRFJ` (Sample J, 2011) and `BCHHRFK` (Sample K, 2012) in the wave-specific datasets `hhrf`.[25]

Additionally to the raking at household level, information on marginal population totals at individual level was used to fit the data to population totals at individual level as well. Again, information on population totals was derived from the German Microcensus. Four variables were used in the raking procedure (see table 6): age, gender, nationality, and household-type[26]. The resulting individual first wave weights are stored in the variables `BBPHRFJ` (Sample J) and `BCPHRFK` (Sample K) in the wave-specific datasets `phrf`.

Researchers interested in using design weights alone are recommended to use the wave-specific variable `design` stored in the dataset `design`. The variable only contains the inverse probability of selection.

[25]For more information on SOEP variables and datasets see Haisken-DeNew/Frick 2005.
[26]Note that information on household type was derived from an analysis of previous SOEP waves instead of using Microcensus data.

7 Characteristics of Combined Cross-Sectional Weights

Table 10 displays characteristics of household weights in each of the three weighting steps of Sample J and Sample K data: design weighting (1), nonresponse adjustment (2), and post-stratification (3).

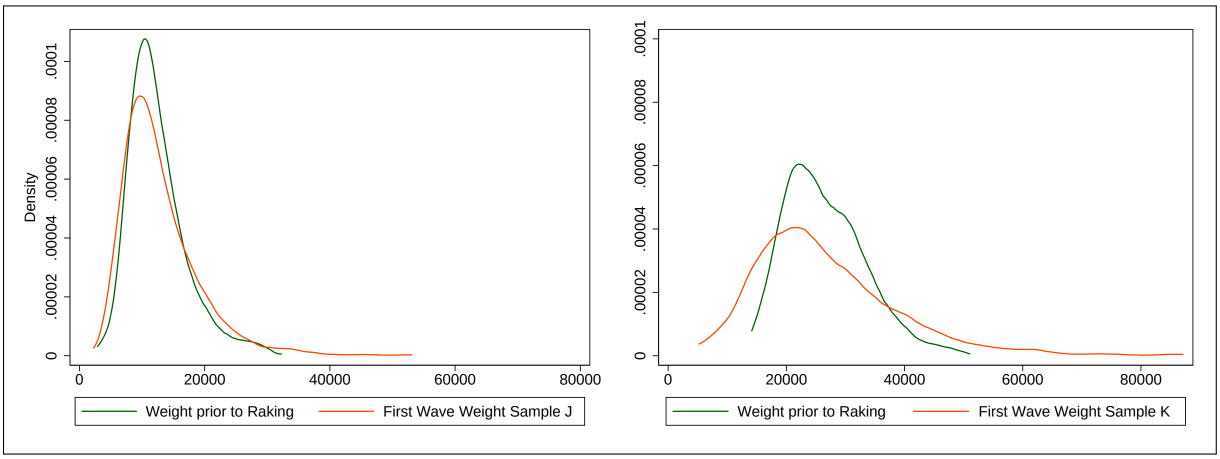
Table 9: Characteristics of Weights during the Weighting Process

		Quantiles								
	Min.	10%	25%	50%	75%	90%	Max.	Mean	SD	n
Sample J										
Design Weight	1549	3066	4132	4251	4427	4719	6264	4171	707	3136*
Design*Nonresponse	2837	8029	9390	11437	14403	18506	32339	12488	4579	3136
Combined First Wave Weight *Raking of Design*NR-Weight*	2222	6927	8798	11436	15576	20572	53147	12906	6167	3136
Sample K										
Design Weight	9179	9179	9179	9179	9179	9179	9179	9179	0	1526*
Design*Nonresponse	14119	18721	21444	25281	30700	35540	51128	26499	6704	1526
Combined First Wave Weight *Raking of Design*NR-Weight*	5189	13823	18123	24202	32715	42660	87183	26681	12279	1526

* Characteristics of design weights refer to the net sample only.

Initial design weights of Sample J and Sample K were grossed-up so that the number of households in the gross sample meets the total number of roughly 40 Mio. households in the target population (residential population of Germany). In comparison to Sample J, we did not assign different sampling probabilities to non/German households in Sample K. Therefore, the design weight for households in Sample K is a constant factor.

Figure 4: Distribution of Weights before and after Raking of Sample J (left) and Sample K (right)



Standard deviations before and after post-stratification reveal that a substantial part of the variance in the combined first wave weights is due to the raking procedure. This is especially true for Sample K data (figure 4). Weights are fairly right-skewed, both prior to raking (green line) and post raking (orange line).

8 Conclusion

This paper analyzed unit nonresponse in the 2011 Refreshment Sample J and the 2012 Refreshment Sample K from the German Socio-Economic Panel study (SOEP). Three goals were pursued. First, the documentation of the sampling design of these samples. Second, variation in nonresponse was to be modeled in order to facilitate a better understanding of properties of both samples. To maximize information on non-participating unit of analysis at the household level, local and regional information was used additionally to fieldwork information on the sampled addresses. Analysis revealed the importance of economical and social background for response rates in a given area, thereby reinforcing past findings. Although model fit at the individual level was low, explanation of observed nonresponse rates at higher levels using geographical data was satisfactory. Third, cross-sectional weights were generated from sampling probabilities, response probabilities, and post-stratification in order to be used in further analysis. Regression design itself as well as post-estimation transformations of weights were guided by the idea of estimating weights within a reasonable range. The use of weights for estimation of parameters using respondents only demonstrated the success of weight generation since bias could be reduced to insignificance using weights in many cases. In addition, exemplary comparison of regression with and without weighting showed the possible impact of the generated nonresponse weights. So far, the nonresponse weights have shown their beneficial impact in estimation. Hopefully, the weights generated here will have the same effect in future work.

A Description of Variables and Expected Effects

Microm-Data

Business intensity (street): Similar to the variable from the field work information but on another level, this variable captures whether the overall picture of a street is dominated by accommodation only or by services, manufacturing or business in general. It has six ascending steps.

Mean age of heads of houses: This variable captures the mean age of heads of households in eight steps, ranging from "under 35 yrs." to "over 65 yrs.". It is compiled at the house cell level. Older people are known to participate less frequently in surveys (Johnson et al. 2006: 711), therefore a negative association with response rates is expected.

Share of families/household structure: This variable displays the dominant structure of households based on the number of household members on the house cell level. Nine steps range from "mainly single person households" to "almost exclusively families w. children". Families with children are expected to be a lot easier to reach and therefore, participation probabilities should be higher in cells dominated by families (Keeter et al. 2006: 768).

Children per household ratio: Quite similar to the last variable, the children per household ratio in nine steps on the house cell level should be associated with higher response probabilities since families with children are more prone to participate (Olson 2006: 746).

Status (socio-economic): This variable is a composite index aggregating education and income of house cell‘s inhabitants. It is coded ascending in nine steps. People with higher status are often believed to be more prone to participation, since they have more often experienced good results in other processes of social exchange (education, career, etc.) and therefore perceive more potential benefit from participation (Durrant/Steele 2009: 375). However, "elites" are sometimes believed to have lower response probabilities. Either way, accounting for status seems necessary.

Share of college graduates: Nine steps ranging from "below 2%" to "over 35%" cover the share of college graduates in the population of a street. Streets with a higher share of college graduates should display higher participation probabilities (Abraham et al. 2006: 694; Singer et al. 1999: 258).

Purchasing power: This index captures a market cell‘s (approximately 470 households) purchasing power in relation to the national average (=100). Purchasing power serves as a proxy for wealth, therefore a positive effect on response is expected (s. above).

Share of immigrants from eastern Europe and Turkey: The population shares of the two biggest groups of immigrants in Germany[27] for each market cell are included in analysis. Controlling for potential differences between both groups seems reasonable, although no specific expectations are held.

[27] See BAMF (2011: 104)

Turnover in accommodation (mobility): This variable captures the turnover in accommodation/housing in nine ascending steps with 5 being the national average. High turnover in a market cell should be associated with shorter time horizons of people. Therefore people might be less likely to participate in multi-wave surveys, especially since giving note on future address change may increase perceived costs of participation. Furthermore, past literature points to the fact that moving may make household members feel less integrated in new environments and reduce participation (Durrant/Steele 2009: 377). While the findings are mixed so far, controlling for mobility seems reasonable.

Balance in accommodation turnover (mobility): On the market cell level, this variable indicates whether turnover in accommodation results in negative (low values) or positive balances (high values). People in potentially less attractive cells with negative balances (that are probably perceived as "disadvantaged areas") should more often plan to move away and therefore be less prone to participate in repeated surveys (see above).

Unemployment: Using a finer distinction of the German 5-digit postal codes (approx. 500 households per cell), unemployment is captured in this variable in 7 steps in relation to the national average. It serves as a proxy for wealth and should therefore be negatively related to response probabilities. Unemployed people are often less in economical structures and may be what Johnson et al. (2006) label "disadvantaged" and others call "isolated" (Durrant/Steele 2009: 375) However, non-contact may be less of a problem with the unemployed, just as with older people.

INKAR

Prices for building ground: Average prices for building ground per square meter on the county level are included as an additional indicator for the wealth and attractiveness of a region and should relate positively to response probabilities.

Average household income: The average disposable household income per person at the county level as an indicator for wealth is included. Positive effects on probabilities are expected.

GDP per capita: Similarly, the GDP per capita ratio is included. It is coded in thousand Euros.

Welfare benefits for renting expenses: County-level data on the height of grants to people entitled to welfare benefits for renting expenses capture the cost of living. Higher grants per person should resemble high costs of living and rather wealthy and more attractive areas. Thus, higher grants should come along with higher probabilities of response.

Share of citizens entitled to welfare benefits: The bigger the share of people entitled to welfare benefits in a county, the less wealthier this county should be. Therefore, a negative effect on response probabilities is assumed.

Medical doctors per inhabitants ratio: Drawing upon the concept of "disadvantaged areas" (Johnson et al. 2006: 707f), higher numbers of doctors in a (more privileged) county should relate positively to response probabilities. It is coded in doctors per 100k inhabitants.

Share of high school and college graduates: It has been shown several times, that higher education of people is associated with higher response probabilities in surveys as education often comes along with a greater sense of civic obligation (e.g. Abraham et al. 2006: 694; Durrant/Steele 2009: 372). Especially high school education should play a vital role and because of its gatekeeper function for college education should produce the biggest differences between individuals. The variables were compiled at the NUTS2-level of EU regions.

Regional Information from the Federal Statistical Office

Electoral turnout: Electoral turnout for the 2009 general parliamentary election was calculated on the community-level in order to capture general participation tendencies. High rates of turnout should relate positively to the affinity of participating in surveys. As has been shown by Keeter et al. (2006: 768) using split-ballot surveys, people that are harder to sample less frequently are registered to vote. On a theoretical level, participation in elections may be related to the same construct as participation in surveys, such as civic obligation.

Vote share for dominant and small parties: As Keeter et al. (ibid.: 768) demonstrate, people with lower response probabilities tend to vote for more ideologically moderate parties. The vote share of the two biggest German parties (SPD; CDU/CSU), which are quite centrist is included to test for this effect. Moreover, the share of small parties (those failing the 5% threshold) is included to account for the possible opposite effect. In addition, modeling of nonresponse in Sample K relies on information of vote shares for other parties generally represented in the German *Bundestag* such as The Left and Alliance '90/The Greens. For example, we expect higher nonresponse rates in areas with a high share of votes for The Greens, as such households tend to be more difficult to be contacted (Schnell 2012: 161).

Share of small and very big flats: The share of differently sized flats in current overall number of flats are included in the model as a proxy for household structures. Small flats are commonly inhabited by single-person-households (which are harder to reach) and bigger flats are more often inhabited by families, that are easier to reach and more prone to participation (Durrant/Steele 2009: 372).

Age structure: Shares of different age groups (7 variables from "18 to 25" in ten year steps to "elderly (65+)") in percent for 2010 (community level) are included as controls. In the past, older people have been shown to be less likely to participate Keeter et al. (2006: 765). Furthermore, age is a important predictor in many social sciences research settings and controlling for it in weight generation seems reasonable.

Further variables

Share of people active in non-profit organizations: The share of people active in non-profit organizations (sporting clubs, churches, community service, etc.) captures general affinity for participation and possibly a sense of civic obligation. People with high affinity for participation possibly do not need high incentives to participate (Durrant/Steele 2009: 378), therefore participation in a survey should be more likely for these people. This variable was compiled at the county level in 2009[28] (Groves et al. 2000: 302f).

Quota of party members: The quota of party members as a part of the general adult population in percent should also be related to general participation affinity. Therefore, positive effects are anticipated.

[28]Data was missing for some counties but for unknown reasons. However, it was extracted manually from Engagementatlas (2009).

References

AAPOR (2011). *Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys. Revised 2011.*

Abraham, Katharine G./Aaron Maitland/Suzanne M. Bianchi (2006). “Nonresponse in the American Time Use Survey”. In: *Public Opinion Quarterly* 70(5), pp. 676–703.

BAMF (2011). *Bundesamt in Zahlen 2010.* URL: `http://www.bamf.de/SharedDocs/Anlagen/DE/Publikationen/Broschueren/bundesamt-in-zahlen-2010.pdf?__blob=publicationFile` (visited on 05/22/2012).

Bethlehem, J./F. Cobben/B. Schouten (2011). *Handbook of Nonresponse in Household Surveys*. Wiley Handbooks in Survey Methodology. Hoboken, NJ: John Wiley & Sons.

Curtin, Richard/Stanley Presser/Eleanor Singer (2005). “Changes in Telephone Survey Nonresponse over the Past Quarter Century”. In: *Public Opinion Quarterly* 69(1), pp. 87–98.

Deming, W.E./F.F. Stephan (1940). “On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known”. In: *Journal of the American Statistical Association* 35, pp. 615–630.

DuMouchel, William H./Greg J. Duncan (1983). “Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples”. In: *Journal of the American Statistical Association* 78(383), pp. 535–543.

Durrant, Gabriele B./Fiona Steele (2009). “Multilevel Modelling of Refusal and Non-Contact in Household Surveys: Evidence from Six UK Government Surveys”. In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 172(2), pp. 361–381.

Engagementatlas (2009). *Engagement Atlas 09 - Daten. Hintergründe. Volkswirtschaftlicher Nutzen. Prognos AG/AMB Generali Holding AG. http://www.deutscher-engagementpreis.de/fileadmin/media/pdf/Sonstige_Download-PDFs/Engagementatlas%202009_Generali.pdf (visited on 11/26/2014)*. (Visited on 02/21/2013).

Gelman, A./Jennifer. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Giraldo, Anna/Gianpiero Dalla Zuanna (2006). “Investigation of a Unit Non-Response Adjustment Procedure: The Case of the Urban Fertility Survey, Italy, 2001-2002”. In: *Population* 61(3), pp. 295–307.

Goebel, Jan/C. Katharina Spieß/Nils R.J. Witte/Susanne Gerstenberg (2007). *Die Verknüpfung des SOEP mit MICROM-Indikatoren: Der MICROM-SOEP Datensatz. SOEP Data Documentation Nr. 26*. Tech. rep. DIW Berlin.

Groves, Robert M. (2006). “Nonresponse Rates and Nonresponse Bias in Household Surveys”. In: *Public Opinion Quarterly* 70(5), pp. 646–675.

Groves, Robert M./Robert B. Cialdini/Mick P. Couper (1992). “Understanding The Decision to Participate in a Survey”. In: *Public Opinion Quarterly* 56(4), pp. 475–495.

Groves, Robert M./Eleanor Singer/Amy Corning (2000). “Leverage-Saliency Theory of Survey Participation: Description and an Illustration”. In: *Public Opinion Quarterly* 64(3), pp. 299–308.

Haisken-DeNew, John P./Joachim R. Frick (2005). *DTC Desktop Companion to the German Socio-Economic Panel (SOEP). Version 8.0 Dec 2005, Updated to Wave 21 (U).*

Horton, Nicholas J/Stuart R Lipsitz (2001). “Multiple Imputation in Practice”. In: *The American Statistician* 55(3), pp. 244–254.

Horvitz, D. G./D. J. Thompson (1952). “A Generalization of Sampling Without Replacement From a Finite Universe”. In: *Journal of the American Statistical Association* 47(260), pp. 663–685.

Humpert, Andreas/Klaus Schneiderheinze (2013). *Linguistic analysis of personal names based on th ”Onomastic process”.* URL: `www.stichproben.de/herunterladen/onomastic_process_HS_GbR.pdf` (visited on 02/20/2014).

INKAR (2011). *Indikatoren und Karten zur Raum- und Stadtentwicklung in Deutschland und in Europa.* URL: `http://www.bbr.bund.de/nn_21272/BBSR/DE/Veroeffentlichungen/INKAR/inkar__node.html` (visited on 05/22/2012).

Johnson, Timothy P./Young IK Cho/Richard T. Campbell/Allyson L. Holbrook (2006). “Using Community-Level Correlates to Evaluate Nonresponse Effects in a Telephone Survey”. In: *Public Opinion Quarterly* 70(5), pp. 704–719.

Kalton, Graham/Ismael Flores-Cervantes (2003). “Weighting Methods”. In: *Journal of Official Statistics* 19(2), pp. 81–97.

Keeter, Scott/Courtney Kennedy/Michael Dimock/Jonathan Best/Peyton Craighill (2006). “Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey”. In: *Public Opinion Quarterly* 70(5), pp. 759–779.

Kim, Jae Kwang/Jay J. Kim (2007). “Nonresponse weighting adjustment using estimated response probability”. In: *Canadian Journal of Statistics* 35(4), pp. 501–514.

Kroh, Martin (2014). *Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984-2013).* Tech. rep. DIW Data Documentation. DIW Berlin.

Lernatlas (2011). *Deutscher Lernatlas.* URL: `http://www.deutscher-lernatlas.de/de/ergebnisse/profile.html` (visited on 05/22/2012).

Lohr, Sharon L. (2010). *Sampling: Design and Analysis.* 2nd ed. Boston, MA: Brooks/Cole.

Lumley, T. (2010). *Complex Surveys: A Guide to Analysis Using R.* Wiley Series in Survey Methodology. Hoboken, NJ: John Wiley & Sons.

McGaw, D./G. Watson (1976). *Political and Social Inquiry.* New York: John Wiley & Sons.

Olson, Kristen (2006). “Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias”. In: *Public Opinion Quarterly* 70(5), pp. 737–758.

Peytchev, Andy/Lisa R. Carley-Baxter/Michele C. Black (2011). “Multiple Sources of Nonobservation Error in Telephone Surveys: Coverage and Nonresponse”. In: *Sociological Methods & Research* 40(1), pp. 138–168.

Regionaldatenbank Deutschland (2012). *Regionaldatenbank Deutschland.* *https://www.regionalstatistik.de/genesis/online/logon*. (Visited on 11/26/2014).

Royston, Patrick (2009). “Multiple imputation of missing values: Further update of ice, wich an emphasis on categorial variables”. In: *The Stata Journal* 9(3), pp. 466–477.

SOEP (2012). *SOEP 2011 – Methodebericht zum Befragungsjahr 2011 (Welle 28) des Sozio-oekonomischen Panels.* TNS Infratest. URL: http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.399480.de/soepmeth_2011.pdf (visited on 05/22/2012).

SOEP (2013). *SOEP 2012 – Methodebericht zum Befragungsjahr 2012 (Welle 29) des Sozio-oekonomischen Panels.* TNS Infratest. URL: http://panel.gsoep.de/soep-docs/surveypapers/diw_ssp0144.pdf (visited on 05/22/2012).

Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data.* Monographs on Statistics and Applied Probability. London: Chapman & Hall.

Schnell, Rainer (2012). *Survey-Interviews. Methoden standardisierter Befragungen.* Berlin: Springer.

Schonlau, Matthias/Martin Kroh/Nicole Watson (2013). “The implementation of cross-sectional weights in household panel surveys”. In: *Statistics Surveys* 7, pp. 37–57.

Singer, Eleanor/Robert M. Groves/Amy D. Corning (1999). “Differential Incentives: Beliefs About Practices, Perceptions of Equity, and Effects on Survey Participation”. In: *The Public Opinion Quarterly* 63(2), pp. 251–260.

Spieß, Martin (2010). “Der Umgang mit fehlenden Werten”. In: *Handbuch der sozialwissenschaftlichen Datenanalyse.* Ed. by Christof Wolf/Henning Best. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 117–142.

Statistisches Bundesamt (2013). *Bevölkerung und Erwerbstätigkeit. Haushalte und Familien - Ergebnisse des Mikrozensus.* Fachserie 1 Reihe 3. Wiesbaden: Statistisches Bundesamt.

Valliant, Richard/Jill A. Dever (2011). “Estimating Propensity Adjustments for Volunteer Web Surveys”. In: *Sociological Methods & Research* 40(1), pp. 105–137.

Van Goor, H./B. Stuiver (1998). “Can Weighting Compensate for Nonresponse Bias in a Dependent Variable? An Evaluation of Weighting Methods to Correct for Substantive Bias in a Mail Survey among Dutch Municipalities”. In: *Social Science Research* 27(4), pp. 481 –499.

White, Halbert (1980). “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity”. In: *Econometrica* 48(4), pp. 817–838.

White, Ian R./Patrick Royston/Angela M. Wood (2011). “Multiple imputation using chained equations: Issues and guidance for practice”. In: *Statistics in Medicine* 30(4), pp. 377–399.

Zaslavsky, Alan M./Lawrence B. Zaborski/Paul D. Cleary (2002). “Factors Affecting Response Rates to the Consumer Assessment of Health Plans Study Survey”. In: *Medical Care* 40(6), pp. 485–499.