The German Socio-Economic Panel study

271

SOEP Survey Papers

Series C – Data Documentation

SOEP — The German Socio-Economic Panel study at DIW Berlin 2015

The 2013 IAB-SOEP Migration Sample (M1): Sampling Design and Weighting Adjustment

Martin Kroh, Simon Kühne, Jan Goebel, Friedrike Preu

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 Documentation (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

Please cite this paper as follows:

Martin Kroh, Simon Kühne, Jan Goebel, Friedrike Preu. 2015. The 2013 IAB-SOEP Migration Sample (M1): Sampling Design and Weighting Adjustment. SOEP Survey Papers 271: Series C. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)

Contact: DIW Berlin
SOEP
Mohrenstr. 58
10117 Berlin

Email: soeppapers@diw.de

Martin Kroh, Simon Kühne, Jan Goebel, Friedrike Preu

The 2013 IAB-SOEP Migration Sample (M1): Sampling Design and Weighting Adjustment

Berlin, 2015

Contents

1 Introduction **7**

2 Target Population and Sampling Frame **10**

3 Sampling Design **13**

3.1 Identification of Target Population Members 14

3.2 Geographic Address Clustering 14

3.3 Sampling of 600 PSUs for Onomastic Analysis 16

3.4 Onomastic Analysis 17

3.5 Sampling of 250 PSUs for Fieldwork 19

3.5.1 Characteristics of the 250 PSUs 20

3.5.2 Test for Sample Bias and Selectivity 22

3.6 Random Walk Algorithm 24

3.7 Characteristics of the Gross Sample 27

4 Results from Fieldwork and Response Rates **29**

4.1 Response Rates by Regions and Country of Origin 30

4.2 Screening of Target Population Members 31

5 Cross-sectional Weighting of Sample M1 **34**

5.1 Design Weighting 35

5.2 Nonresponse Weighting Adjustment 37

5.2.1 Modeling Nonresponse 38

5.2.2 Correlates of Nonresponse: Data Sources 39

5.2.3 Multiple Imputation and Data Coding 45

5.2.4 Results and Calculation of Nonresponse Weights 46

5.3 Poststratification and Raking 53

6 Characteristics of Combined Cross-Sectional Weights **57**

A Description of Variables and Expected Effects **59**

List of Tables

1 Distribution of PSUs by Type of County 16

2 Regions of Origin by Citizenship and Onomastic Classification 18

3 Composition of the Members in the 250 PSUs by Country of Origin and Generation 21

4 Comparison of Socio-economic Characteristics between Different (Sub)Samples of PSUs 23

5 Address Sampling Factors by Country of Origin 25

6 Composition of Population and Samples by Nationality/Country of Origin 28

7 Final Results from Fieldwork (Households) 29

8 Response Rates by Migrant Groups 31

9 Household Screenouts by Migrant Groups 33

10 List of Variables Used in Analysis of Nonresponse of Sample M1 – I 43

11 List of Variables Used in Analysis of Nonresponse of Sample M1 – II . . . 44

12 Share of Missings in Sample M1 for Different Variables by *Länder* 45

13 Fit Values for Estimated Models 47

14 Comparison of Estimated and Actual Response Rates by Sample Point . . 50

15 Characteristics of Raw Estimated Nonresponse Weights 50

16 Comparison of Weighted and Unweighted Estimates and Reduction of Bias 52

17 Population Characteristics Used in the SOEP Raking Procedure at Household Level 54

18 Comparison of Strata Composition Prior to Post-Stratification (Household Representatives) 55

19 Population Characteristics Used in the SOEP Raking Procedure at Individual Level 56

20 Characteristics of Weights During the Weighting Process 57

List of Figures

1 Steps of the Sampling Procedure 13

2 Number of PSUs by County - 6,725 PSUs 15

3 Visual Representation of Address Sampling in a Given PSU 27

4 Response Rates by *Länder* and Counties 30

5 Screening and Identification of Target Population Anchors 31

6 Coefficients and Confidence Intervals for the Estimated Reduced Model . . 48

7 Distribution of Weights at all Three Weighting-Steps 58

1 Introduction

The 2013 *IAB-SOEP Migration Sample (M1)* is a household survey conducted jointly by the Institute for Employment Research (IAB) in Nuremberg and the German Socio-Economic Panel (SOEP) at DIW Berlin. The first wave of the survey was carried out in 2013, where a direct interview was sought with 4,964 persons in 2,723 households. Brücker et al. (2014) provide a general overview of the survey design as well as its contents. The present paper describes in more detail the sampling and weighting strategy for this sample. We first describe the target population and the sampling frame, namely employment registers of the Federal Employment Agency (BA). We then elaborate the two-stage stratified sampling design. The following section documents the non-response analysis in the gross sample. Finally, based on selection probabilities (design-based weights), response-probabilities (model-based weights) and post-stratification, we document the cross-sectional weights of the IAB-SOEP Migration Sample (M1) (`bdhhrfm` for households and `bdphrfm` for persons).

About IAB and SOEP

The **Institute for Employment Research ("Institut für Arbeitsmarkt- und Berufsforschung", IAB)** is an independent institute of the Federal Employment Agency in Nuremberg, Germany. The IAB conducts labour market and occupational research in Germany. Research deals with topics like labour market policy or social inequality. Besides, the IAB does research in the field of statistical methods and survey methodology. The Research Data Centre (FDZ) of the Federal Employment Agency is associated at IAB where a wide variety of data for research purposes is offered.[a]

The **German Socio-economic Panel (SOEP)** is a longitudinal survey of private households in Germany produced by the German Institute for Economic Research (DIW Berlin) and is annually conducted since 1984. Since 2002, the SOEP receives continued funding through the Joint Science Conference by the Federal Government and the State of Berlin. Before that, SOEP was primarily funded through the German National Science Foundation (DFG). The survey provides information on various topics such as household composition, employment, health and attitudes. In 2012 a total of 20.806 persons in 12.322 households were interviewed (TNS Infratest 2013).

[a] For more information on the IAB visit the webpage: `http://www.iab.de/en/ueberblick.aspx`.

The Changing Structure of the Migrant Population in Germany

Since the 1960s, immigration and the integration of migrants have become more and more important topics in German politics and society. Moreover, various scientific disciplines investigating the causes and consequences of immigration have been constantly growing over the last decades. Political initiatives such as the "Integrationsgipfel" or the monitoring of integration of the Bundesregierung ("Integrationsmonitoring") reflect both the importance and necessity of reliable and valid data for political decision-making processes.

In the German context, immigrants and their descendants are usually referred to as the "population with a migration background". The German population with a migration background consists of different subgroups and can be understood as a rather heterogeneous population. According to the definition of the Federal Statistical Office of Germany ("Statistisches Bundesamt", FSO) the migrant population consist of three main subgroups (FSO 2013: 5f):

1. Persons who immigrated into the territory of today's Federal Republic of Germany after 1949;
2. Foreigners born in Germany;
3. Persons born in Germany with the German nationality who have at least one parent that (1) either immigrated to Germany after 1949 or (2) was born as a foreigner in Germany.

Even though the SOEP included samples of migrants from its very beginning - Sample B (1984) so-called "Guest Workers" from Southern Europe and Sample D (1994) so-called "Late Repatriates"[1] from Eastern Europe as well as migrant boosts of the cross-sectional refreshment samples F (2000), H (2006), and J (2010), the constantly changing structure of the migrant population in Germany raises a challenge to the coverage of this population in an ongoing longitudinal survey.

For instance, the share of second-generation migrants who were born in Germany and are often in the possession of the German nationality has grown since the 1990's. By now, they represent one third of the population with migration background (FSO 2013). Furthermore, over the last years, the immigration from traditional recruitment countries of labor migration in Southern Europe was gradually replaced by immigration from middle and Eastern Europe in the 1990s. This development is in part a result of the collapse of the Soviet Union in 1989/90. In the years following the former many late repatriates and Jewish quote refugees migrated from the successor states of the former Soviet Union to

[1] Late repatriates are German nationals according to Article 116 Basic Law for the Federal Republic of Germany and thus received the German nationality immediately after their arrival. Late Repatriates are referred to as "Ethnic Germans" as well.

Germany. Moreover, the eastern enlargement of the European Union (EU) in 2004 and 2007 and the following regulations concerning freedom of movement and family reunion lead to more immigration from Central and Eastern Europe, with Poland being the primary country of origin (BAMF 2013).

Hence, there is a need for longitudinal data particularly on second-generation migrants and recent immigration to Germany.[2] The Institute for Employment Research (IAB) and the Socio-Economic Panel Study (SOEP) aimed to contribute to the closure of this data gap by creating a representative, innovative, and sustainable database for the (inter)national research community. The collaboration of the two institutions offers the possibility of linking the administrative data of the Federal Employment Agencies with additional information collected as part of the SOEP.[3] The resulting database will contain detailed information on many areas such as labor market situation, working and educational biographies in the countries of origin, experiences with immigration and integration, family situation, personal characteristics, and aspirations.

Besides contributing to the closure of the described data gap, the new sample will be integrated in the set of existing SOEP samples and thereby double the number of migrants in the SOEP. More specifically, the 2013 IAB-SOEP Migration Sample (M1) represents partly a refreshment sample of groups, such as the second-generation migrants, that are already part of the SOEP, however, not in sufficient numbers. Moreover, Sample M1 is an enlargement sample of SOEP that covers groups of recent migrants to Germany that arrived after sampling of the previous migrant samples (Kroh et al. 2015). While Sample B covers immigration to Germany until 1983, Sample D covers immigration between 1984 and 1994, the 2013 IAB-SOEP Migration Sample (M1) immigration since 1995.

[2]The German Microcensus includes questions concerning the respondents' migration background (FSO 2010) and therefore enables analyses beyond a simple distinction by citizenship. However, as the Microcensus constitutes a rotating short term panel, no long-term processes can be analyzed. Furthermore, the questionnaire of Microcensus is limited to general information on, for instance, occupation, education, and living conditions and is only limited in covering migrant-specific questions and questions of social integration and well-being. Longitudinal studies such as the SOEP can in principle survey these missing characteristics, but they can not comprehend the employment histories as precisely as register data.

[3]The "Record Linkage" of the resulting survey data to the administrative data of the FEA and IAB opens up new possibilities for an in-depth analyses of migration and integration processes. Both survey data and administrative process-data on the individual level could therefore be analyzed simultaneously. Selectivity in consenting or non-consenting to record linkage could be analyzed and corrected on the basis of the detailed IEB information about all the individuals in the gross sample (Eisnecker et al. 2015). In addition, informed consent is obtained in an experimental design that assures the possibility to analyze and control for the impact of record linkage (in particular, the obtaining of consent) on long-term participation rates. See Schnell (2013) for a brief overview on record linkage processes of survey and administrative data.

2 Target Population and Sampling Frame

In the case of the 2013 IAB-SOEP Migration Sample (M1), the target population consists of immigrants who arrived in Germany since 1995[4] as well as second-generation migrants. Both groups constitute a subset of the population group with migration background (see section 1). Second-generation migrants are individuals that have at least one parent that immigrated into the Germany after 1949. They can have the nationality of their parents, the German nationality, or both. Immigrants or first-generation migrants are defined as individuals who were born in a foreign country and immigrated into Germany, in the case of Sample M1: since 1995. Many are still in possession of the nationality of their country of origin. They have the German nationality if they were naturalized or if they came to Germany as late repatriates. We will speak of “members of the target population” if we want to make statements about both groups.

Germany lacks a centralized register of persons with migration background. Although local registers of municipalities (Melderegister) in principle record the nationality of their inhabitants, sampling of naturalized persons with migration background and sampling of certain immigration-cohorts is practically impossible (Salentin 1999).[5] Alternative sampling strategies, also employed in the context of the SOEP in the past, use either a large number of screening interviews, by telephone for instance (Sample D), or onomastic procedures relying on information of addresses (Samples F, H, and J).

Telephone-based screening as a source of sampling migrants became increasingly problematic in recent years. The growing number of persons and households which either do not have a land-line or are not registered in the phone directory constitutes a source of sample selectivity. This has been shown to be especially true for the migrant populations, who are frequently not in possession of a land-line (Lipps/Kissau 2012). Also, due to the comparatively small proportion of target population members within the sampling frame, screening interviews would be inefficient and expensive. The same holds for onomastic procedures drawing on visible family names near door bells. While we use the procedure of onomastics to extend the coverage of our final sampling frame (see section 3.4), we

[4]The previous migrant samples covered migration until 1983 (Sample B in 1984) and between 1984 and 1994 (Sample D in 1994/5).

[5]One exception are immigrants at their first registered address in Germany (Diehl 2007). A central register containing migrants is the “Central Register of Foreigners” (Ausländerzentralregister) provided by the Federal Office for Migration and Refugees. Access to the data is restricted to the Federal Institute of Population Research.

refrained from using only this method for obtaining a sampling frame. Not least, since the method is unable to identify years of immigration and concepts such as first- and second-generation migrants.

An alternative sampling frame, which has not been used in the context of migration studies in Germany before, is the register data of the Federal Employment Agency (FEA), the so-called Integrated Employment Biographies (IEB). The IEB covers employees, unemployed persons, job seekers, recipients of mean-tested benefits (unemployment benefit II) and participants in active labor market programs on a daily basis from 1975 onward. The IEB is available since 2004 and provides information from as far back as 1990 (Oberschachtsiek et al. 2009). The IEB can thus be understood as the result of a merging procedure of different process-produced databases (Oberschachtsiek et al. 2009, Jacobebbinghaus/Seth 2007):

1. Employment: IAB Employee History (BeH)
2. Unemployment: Benefit Recipient History (LeH)
3. Active labor market policies: Participants-in-Measures History File (MTH)
4. Job search: Applicant Pool Data (BewA).

Data is collected on employees in Germany who are subject to social security contributions, which describes almost all private sector employment. Employers are requested to submit information on starting and ending dates of all their employees' job spells as well as total earnings received (censored at the maximum taxable earnings level) on an annual basis. Data is stored in spells that are linked to persons. The spells come along with a variety of socio-demographic variables such as gender, age, or nationality as well as spatial information such as addresses, municipality, and region type. In total, the IEB contains 83,521,672 individuals with 1,894,018,836 spells.[6] Furthermore, information on unemployment spells, benefit receipt, participation in active labor market policies, and job-search status are directly matched from the different sources of the social security system to form a complete picture of the labour market history of individuals.

Three advantages come with the IEB as a sampling frame. First, a practical advantage is that the IEB is a centralized sampling frame whereas register offices in Germany work

[6]Key date December 31st 2011.

at the local level and national sampling requires collaboration with each of the sampled municipalities. Second, the wealth of information on the labor market participation of individuals, their wages, as well as information on their employers enables researchers to model non-response processes more fully than is possible with many alternative sampling frames. The (model-based) weighting of the data obtained in the IAB-SOEP Migration Sample (M1) – for instance, on wages – thus corrects for any deviation in registered wages in the IEB between the gross and the net sample. Third, the IEB sampling frame allows the linkage of survey and register data in subsequent research projects.

The IEB has some disadvantages as well. Although the database represents a great share of the target population, some groups are not covered (on "undercoverage" in the IEB, see Jacobebbinghaus/Seth 2007). Public sector employees are only covered if they are obliged to pay social security contributions; civil servants who are not covered by the social security insurance system (so-called "Beamte") are not. Also, self-employed people who have never held a job that is subject to social security contributions and have never received unemployment benefits or attended an active labor market policy measure are not covered in the IEB. All in all, the IEB covers about 80 % of the cross-section of the German labor force (Jacobebbinghaus/Seth 2007: 336). An estimation based on the SOEP and the German Microcensus has shown that by choosing the IEB as a sampling frame, 5 to 8 percent of the target population is excluded. However, the excluded groups, such as the self-employed, students, and refugees, may enter the survey as household members of anchor persons (see below), but possibly less than proportionally.

3 Sampling Design

In this section, we describe the multi-stage sampling design of the migrant sample, which is graphically represented by Figure 1. As described above, we use the IEB as the sampling frame. The first step in the sampling process was the identification of likely target population members which included a preliminary classification of individuals on the basis of nationality and policy measures designed for migrants. Second, individual addresses were geocoded and geographically clustered to primary 6,725 "primary sampling units" (PSUs), respectively "sample points". A number of 600 PSUs that comprised a total of about 1 Mio. individuals was then sampled to be included in an onomastic procedure of first and last names of persons living in these PSUs. Based on this additional piece of information that allowed us to identify migrants with German nationality, stratified random sampling was used in order to select a smaller number of 250 sample points for the fieldwork. Finally, a total of 80 households were sampled in each PSU resulting in 20,000 records of the gross sample by implementing a simulated random walk procedure with disproportional sampling probabilities for different migrant groups.

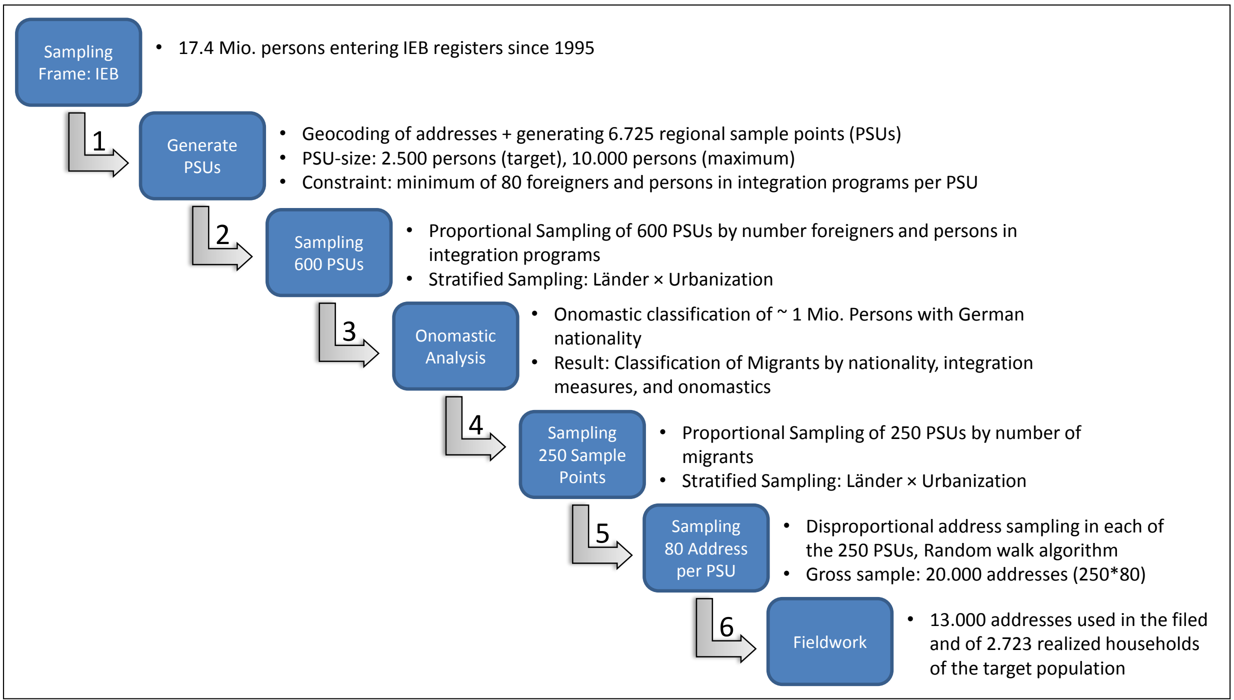

Figure 1: Steps of the Sampling Procedure

3.1 Identification of Target Population Members

The first step in the sampling process was a preliminary identification of target population members. As we were interested in individuals who immigrated since 1995 as well as second-generation migrants (descendants of migrants), we reduced the total IEB database – comprising about 67 Mio. individual data entries (key date: December 31, 2011) – by using information on first appearance in the register. All individuals whose first appearance was prior to the key date of January 1st 1995 were excluded. Moreover, we excluded records from the IEB if the last information dated before January 1st 2007. Pretests revealed that these outdated records were plagued by incorrect addresses in about 50 percent of all cases.

In the remaining records, we implemented a preliminary coding of likely members of the target population on the basis of a foreign, i.e. non-German citizenship (today or in the past) as well as individuals who ever took part in measures of the FEA specifically designed for persons with migration background (e.g. language classes). We use this information on likely target population members in the next step for constructing regional sampling points and also for the sampling of these points which were to be used in fieldwork.

Furthermore, we exclude individuals preliminary classified as second generation migrants and born prior to 1977, since these cohorts are infrequent in the second generation of migrants to Germany and incidences in the IEB are likely to reflect misclassifications. Morever, the restriction of first employment and transfer spells since 1995 implies selection on late labor market entries in cohorts born before the mid 1970s.

The restrictions reduced the database to about 17.4 Mio. individual records from which about 23.6 percent (4.1 Mio.) were preliminary classified as either first or second generation migrants.

3.2 Geographic Address Clustering

Available address information on records was used to create spatial distinct sample points based on a clustering of geographically nearby addresses.[7] Geocoding prior to the clustering was successful for about 90 % of all 17.4 million addresses. While the proportion

[7] Clustering was based on geocoded addresses, i.e. a geographical position in latitude and longitude. Geocoding and clustering was conducted by the team of the German Record Linkage Center (`www.record-linkage.de`). The German Record Linkage Center is associated at IAB and University of Duisburg-Essen.

of addresses for which geocoding was not feasible differed greatly by county (min.: 2.8 %, max.: 31.6 %), it did differ only slightly by population group (migration background: 9.6 %, no migration background: 10.1 %).

Three restrictions were implemented in this process. First, sample points ought to include a minimum of 2,500 and a maximum of 10,000 persons. Second, sample points had to cover a minimum of 80 members of our target population based on our preliminary classification of migrants.[8] These thresholds were set to ensure sufficiently large numbers of target households in each PSU for further stages of the sampling procedure. Third, to ease the face-to-face fieldwork, we restrict geographic spread of PSUs by not allowing them to cross the official administrative borders of the 402 counties in Germany. This procedure resulted in a total number of 6,725 sample points.

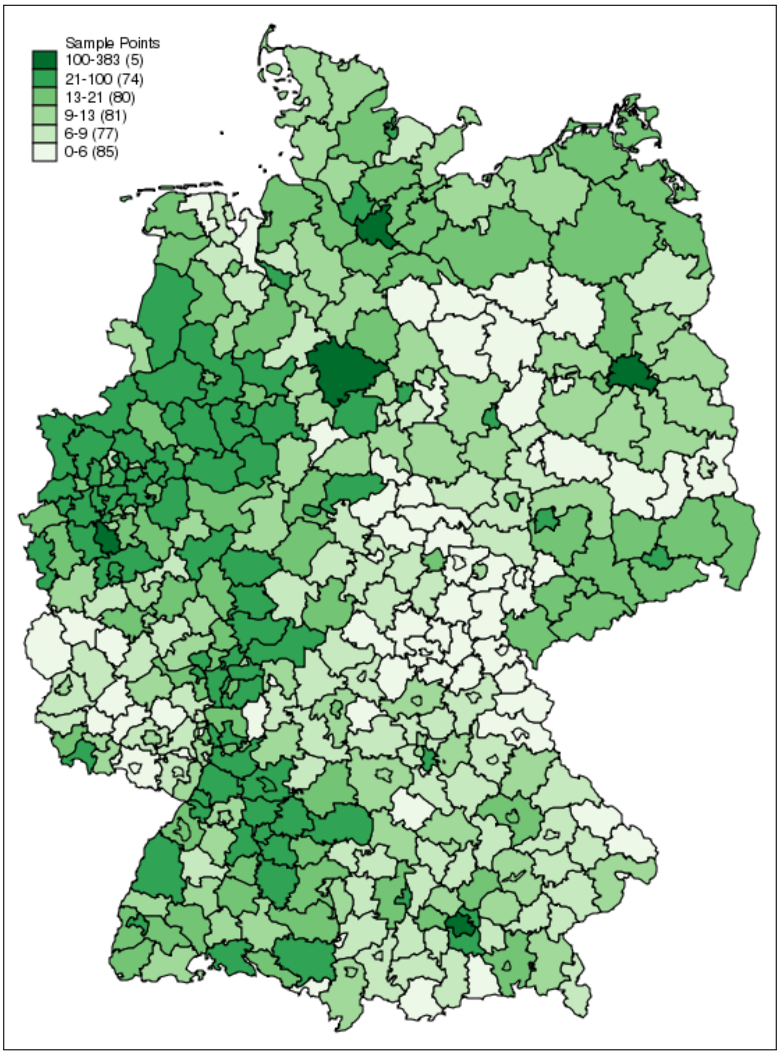

Figure 2: Number of PSUs by County - 6,725 PSUs

[8]Due to conflicting restrictions, some PSUs contained less than 80 individuals being classified as migrants. These "uncompleted" sample points were later assigned to other nearby sample points.

Table 1: Distribution of PSUs by Type of County

Counties (N)	Min. number of PSUs	Median (Mean)	Max. Number of PSUs
Urban Agglomeration (136)	3	16.5 (27.9)	383
Urbanized (191)	3	10 (12.2)	35
Rural (75)	2	7 (8.0)	26

Figure 2 displays the number of PSUs by county, which ranges from 2 (Suhl and Zweibruecken) to 383 (Berlin). As one would expect, the number of PSUs per county is higher in urban agglomerations and urbanized areas than in rural ones (see Table 1).

3.3 Sampling of 600 PSUs for Onomastic Analysis

As mentioned above, we used an onomastic procedure (see section 3.4) to improve the classification of descendants of immigrants who hold German citizenship and who could have not been identified relying exclusively on information available from the IEB records. These members of the target population are part of the much larger group of 13.3 Mio. persons who were always coded as German citizens. Due to practical and economical limitations, the onomastic procedure was conducted for about 1 Mio. out of these 13.3. Mio. persons. A subsample of 600 from the total of 6,725 PSUs roughly covers this figure of 1 Mio. persons with these characteristics.

PSU selection was based on a stratified random sampling by the 16 German states (*Länder*) and a classification of the level of urbanization of counties into 9 categories [9]. Due to the presence of empty cells, the number of 144 (16×9) possible cells was reduced to 71.[10] The number of PSUs to be sampled in each stratum was proportional to the number of individuals classified preliminarily as migrants in a given stratum.

Strata are “mutually exclusive”, i.e. every PSU is assigned to one specific stratum, h, only. In our given case, $n_h^{(1)}$ denotes the number of PSUs to be drawn from the total number of PSUs in each stratum, N_h. The number of selected PSUs per stratum was proportional

[9]There were nine different county types ranging from ”rural area with low population density” to “agglomeration area”. We made use of the settlement structure classifications developed by the Federal Office for Building and Spatial Planning. For more information visit `http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/SiedlungsstrukturelleGebietstypen/Kreistypen/kreistypen.html`.

[10]Either because certain region types did not exist in some states or region types were combined due to small number of cases.

to the number of individuals classified preliminarily as migrants in a respective stratum, $M_h^{(1)}$, relative to the overall number, $M^{(1)}$, which is 4.1 Mio. Since geocoding was not always successful and the rate of successful geocoding varied by region, we use all data with and without geocoding for calculating $M_h^{(1)}$:

$$n_h^{(1)} = n^{(1)} \times \frac{M_h^{(1)}}{M^{(1)}} \tag{1}$$

with

$$n^{(1)} = \sum_{h=1}^{71} n_h^{(1)} = 600 \ , \ \ N^{(1)} = \sum_{h=1}^{71} N_h^{(1)} = 6,725 \ \text{ and } M^{(1)} = \sum_{h=1}^{71} M_h^{(1)} = 4,100,000 \tag{2}$$

Within each stratum, random sampling of n_h PSUs was proportional to the number of migrants, given the tentative identification of the target population. Since the construction of PSUs required geocoding of addresses, the selection probability of sampling point k in stratum h, $\pi_{k,h}$, depends on the number of geocoded individuals classified preliminarily as migrants multiplied by the inverse of the county-level rate of geocoded adresses, $M_{k,h}^{(1)*}$:

$$\pi_{k,h}^{(1)} = \frac{M_{k,h}^{(1)*}}{M_h^{(1)*}} \tag{3}$$

with

$$M_h^{(1)*} = \sum_{k=1}^{K_h} M_{k,h}^{(1)*} \ \text{ and } M^{(1)*} = \sum_{h=1}^{71} M_h^{(1)*}. \tag{4}$$

3.4 Onomastic Analysis

So far, the identification of the targeted migrant group was based exclusively on information on citizenship and attendance to FEA measures derived from the IEB. However, this approach excluded members of the target group who were always registered as German citizens and never took part in a FEA measure. This limitation particularly affects the representation of children of immigrants in the sample. To improve their coverage, all persons listed in the 600 sampled PSUs with German citizenship (now and in the past) were coded into categories of regions of origin based on their given names as well as their family names (Humpert/Schneiderheinze 2013). In the "onomastic" procedure, names are compared to large databases containing country and ethnic/origin specific lists of names (see Schnell et al. 2013: 4). Individuals are then assigned to a region of origin based on a probabilistic matching procedure.[11]

Several previous studies used onomastics to identify migrants in sampling frames that provide names, but not sufficient information on place of birth, for instance (e.g., the SOEP Samples F and J, but also, for instance, the PIONEUR project, see Santacreu Fernandez et al. 2006). Clearly, the method is not without problems and mis-classification is a frequent one (e.g., Santacreu Fernandez et al. 2006, Liebau 2011). Moreover, matching rates between the classification of the name and the actual country of origin vary systematically across countries/regions of origin. As onomastics is based on languages, the categories represent the linguistic origin of a name and not essentially the nationality of the person. Obviously, migrants from German speaking countries and Late Repatriates are difficult to distinguish from natives. Also, classification into different sending countries can be a problem particularly for languages used in many parts of the world, such as Spanish and Arabic. This is why Table 2 groups countries of origin to broader categories of regions. These 11 regions also reflect the major sending countries of migrants to Germany.

Table 2: Regions of Origin by Citizenship and Onomastic Classification

Origin	Citizenship Only	+ Onomastic	Citizenship Only	+ Onomastic
	Frequency		Percent	
Italy	19,923	28,114	1.27	1.80
Spain	4,056	5,546	0.26	0.35
Greece	11,113	12,714	0.71	0.81
Turkey	111,211	133,410	7.11	8.53
Former Yugoslavia	42,816	49,143	2.74	3.14
Late Repatriates	/	60,376	/	3.86
Poland	28,886	43,154	1.85	2.76
Rumania	11,890	12,976	0.76	0.83
Former CIS	50,012	53,371	3.20	3.41
Arabic Countries	28,110	56,424	1,80	3.61
Rest of World	129,383	181.945	8,27	11.63
Subtotal	437,400	637,173	27.95	40.72
Germany	1,127,419	927,646	72.05	59.28
Total	1,564,819	1,564,819	100.00	100.00

The first column reports the region of origin for the 1.56 Mio. persons in the sampled

[11] Another name-based technique to identify migrants and their country of origin was recently proposed by Schnell et al. (2013). The procedure is based on n-grams – substrings of text of length n – of names in order to estimate country-of-origin related probabilities.

600 PSUs based on the information on nationality available in the IEB. A total number of 1.13 Mio. persons who were always registered as German citizens made part of the onomastic procedure. The method classified about 200,000 of these individuals as migrants respectively their descendents. The second column of Table 2 lists the cumulative number of persons by region of origin of the 437,400 persons with non-German nationality (at present or the past) *and* the 199,773 persons classified as migrants by the onomastic procedure. While we process in the following steps the total of 637,173 records of presumed migrants, we exclude the 927,646 persons from sampling with (always) German citizenship, never participating in migrant training as well as integration courses, and having a German name.

3.5 Sampling of 250 PSUs for Fieldwork

In the next step, we sampled 250 of the 600 sample points for fieldwork. Since the sampling probabilities of the previous step used primarily information on nationality, sampling probabilities in this step account for regional variation in the number of additional identified migrants of the onomastics. Due to the much smaller number of remaining PSUs, we collapsed the former 71 strata of states and the country typology of urbanization to 52.

$$n_h^{(2)} = n^{(2)} \times \frac{M_h^{(2)}}{M^{(2)}} \tag{5}$$

with

$$n^{(2)} = \sum_{h=1}^{52} n_h^{(2)} = 250 \ , \ \ N^{(2)} = \sum_{h=1}^{52} N_h^{(2)} = 600 \ \text{ and } M^{(2)} = \sum_{h=1}^{52} M_h^{(2)} = 637,173 \tag{6}$$

The sampling probability in the second step, $\pi_{k,h}^{(2)}$, depends on the number of persons of the target population per PSU. Since the first step used information on nationality already, we divide the sampling probability of the second step by the first-step probability. The sampling probability in the second step thus relies on the number of additionally identified persons of the target population in the onomastic procedure:

$$\pi_{k,h}^{(2)} = \frac{M_{k,h}^{(2)*}}{M_h^{(2)*}} \times \frac{1}{\pi_{k,h}^{(1)}} \tag{7}$$

with

$$M_h^{(2)*} = \sum_{k=1}^{K_h} M_{k,h}^{(2)*} \text{ and } M^{(2)*} = \sum_{h=1}^{52} M_h^{(2)*}. \tag{8}$$

The combination of first and second step sampling probability is then given by

$$\pi = \pi^{(1)} \times \pi^{(2)} = \pi^{(1)} \times (\frac{M_{k,h}^{(2)*}}{M_h^{(2)*}} \times \frac{1}{\pi^{(1)}}) = \frac{M_{k,h}^{(2)*}}{M_h^{(2)*}} \tag{9}$$

In other words, the combination of both steps results in equal selection probabilities of persons of the target population in Germany, including the number of persons with and without German nationality.

3.5.1 Characteristics of the 250 PSUs

Table 3 displays the number of persons in the 250 sampled PSUs by country of origin and generation. In order to distinguish between first- and second-generation migrants, we use a tentative classification of immigrants and descendants of immigrants drawing on available information in the IEB. Most of them are based upon the age of the person at the time of his/her first IEB spell as well as his/her educational biography. For example, all persons who were under 24 when they first entered the IEB and for whom vocational training was reported were classified as second-generation migrants. Also, we distinguish only between first- and second-generation migrants in those countries of origin with a longer history of migration to Germany, i.e. the former recruitment countries of labor migration from the 1950s through the 1970s. In some cases, a classification was not possible due to the lack of clear and lucid information. These individuals were assigned to the special category of "non differentiable (n.d.)".

Table 3: Composition of the Members in the 250 PSUs by Country of Origin and Generation

Origin	Generation	n	%
Italy	*1*	*2,909*	*26.5*
	2	*6,699*	*61.1*
	n.d.	*1,353*	*12.3*
	all	10,961	4.3
Spain/Greece	*1*	*2,660*	*36.8*
	2	*3,428*	*47.4*
	n.d.	*1,148*	*15.9*
	all	7,236	2.9
Turkey	*1*	*17,425*	*32.8*
	2	*28,957*	*54.5*
	n.d.	*6,794*	*12.8*
	all	53,176	21.0
Former Yugoslavia	*1*	*8,127*	*43.4*
	2	*8,394*	*44.8*
	n.d.	*2,197*	*11.7*
	all	18,718	7.4
Late Repatriates	all	31,252	12.4
Poland	all	16,375	6.5
Rumania	all	4,480	1.8
Former CIS	all	18,276	7.2
Arabic Countries	all	22,597	8.9
Rest of the World	all	69,547	27.5
Total		252,618	100.0

As can bee seen from Table 3, Turkish immigrants and their descendants represent the largest migrant group (21%) in the 250 selected sample points. The classification by generation in migrants from Italy, Spain/Greece, Turkey and Former Yugoslavia reveals further differences between countries. Although second-generation migrants represent the largest group within all of these countries, the Turkish and Italian groups have a higher proportion of second-generation migrants (61.1% and 54.5%) compared to migrants from Spain/Greece and Former Yugoslavia (47.4% and 44.8%). The proportion of individuals for whom a classification of generation was not possible ("non differentiated", n.d.) is comparatively constant across groups (around 12.5%), however, classification was less successful for migrants from Spain and Greece (15.9%). Finally, the large category "'Rest of the World"', representing the 27.5% of the total sample, suggests that migration to Germany is increasingly diverse.

3.5.2 Test for Sample Bias and Selectivity

Sampling error may produce differences in the properties of the sample and the underlying population. To test for a potential bias, we compared various characteristics of PSUs between (1) the 6,725-PSUs and 600-PSUs as well as (2) the 600-PSUs and the 250-PSUs. For this purpose, we conducted t-tests for differences. Table 4 displays the results drawing on county-level statistics on various socio-economic indicators such as education, economy & housing, labor market, demography as well as nationality provided by the Federal Statistical Office (see INKAR 2011).

A direct comparison of estimates may lead to misinterpretation given that the differences between the levels can be partly explained by the selection probabilities focussing on migrants, not on the cross-sectional population. Since it is not possible to distinguish between the part of the difference resulting from the selection probabilities and the biased part just by comparing unweighted or "raw" values of the 6,725 PSUs (column I in Table 4) with the values of the 600 PSUs (column III), it is necessary to weigh the "raw"-values with the respective selection probabilities in the first step in order to balance the differences. Afterwards, the difference between these weighted values (column II) and the values of the 600 PSUs (III) can be tested for significance. The same is true for the comparison between the 600-PSU level and the 250-PSU level. The "raw"-values of the 600 PSUs (column III) have to be weighted by the selection probability in the second step prior to significance testing. Afterwards, the difference between the weighted values (column V) and the values of the 250 PSUs (column VI) can be tested for significance.

The results of the t-tests (columns IV & VII) do not indicate any statistically significant differences with regard to the variables used, neither between the 6,725-PSU level and the 600-PSU level, nor between the 600-PSU level and the 250-PSU level. Thus, both samples do not seem to be subject to selectivity caused by the sampling procedure.

Table 4: Comparison of Socio-economic Characteristics between Different (Sub)Samples of PSUs

Variable	Mean			t-Test: p-value	Mean		t-Test: p-value
	I	II	III	IV	V	VI	VII
	6,725 PSUs Unweighted	6,725 PSUS Weighted	600 PSUs Unweighted	II-III	600 PSUs Weighted	250 PSUs Unweighted	V-VI
Education							
Number of University Students per 10 Residents	.3169	.3621	.3706	.6504	.3584	.3511	.7918
Prop. of Highly Qualified Graduates ("Allg. Hochschulreife")	.3144	.3141	.3146	.8688	.3159	.3121	.4565
Economy & Housing							
GDP per Capita in Thousand Euros	31.0322	34.2231	34.4566	.6809	33.7127	33.6524	.9448
Tax Revenue per Capita & per Year in Euros	647.0132	705.7439	708.3007	.7912	694.8675	692.5512	.8777
Balance of Migration	-.2008	-.2860	-.2860	.9995	-.2531	-.3168	.5689
Population Density	2592.8420	2956.2880	2966.2480	.8644	2890.6930	2881.1720	.9169
Building Land Prices per sqm. in Euros	197.1346	245.9955	248.5884	.7654	237.8046	239.6996	.8872
Prop. of Single-person Households	.3997	.4097	.4100	.9249	.4078	.4078	.9997
Labor Market							
Unemployment Rate	.0633	.0615	.615	.9895	.0620	.0619	.9787
Prop. of Unemployed Immigrants	.1772	.2134	.2143	.8105	.2073	.2080	.9060
Unemployment Benefits 1: Average Payment per Capita & per Month in Euros	775.7701	792.211	791.6972	.9152	788.7499	788.3008	.9051
Unemployment Benefits II: Proportion of Receivers Relative to the Population Under 65	.1072	.1065	.1066	.9629	.1071	.1072	.9741
Proportion of Low-qualified Workers	.3188	.3271	.3279	.6464	.3270	.3272	.9246
Demographics							
Prop. of Persons Aged 18-25	.0839	.0838	.0840	.5632	.0840	.0838	.7357
Prop. of Persons Aged > 65	.2041	.2005	.2000	.4514	.2009	.2009	.9660
Proportion of Foreigners	.0960	.1154	.1163	.6673	.1123	.1125	.9546
Number of IEB Immigrants	670.7134	878.374	889.8184	.4809	851.1713	862.8880	.6440
Share of Nationality							
Germany: no Migration Background (IEB)	.7614	.6926	.6880	.3942	.7018	.6991	.7632
Germany: Migration Background (IEB)	.0258	.0292	.0296	.7439	.0293	.0290	.8902
Italy	.0098	.0130	.0129	.8739	.0122	.0121	.9307
Spain	.0020	.0024	.0026	.1159	.0025	.0025	.8763
Greece	.0050	.0070	.0072	.4553	.0067	.0071	.5605
Turkey	.0500	.0710	.0716	.7975	.0677	.0687	.7672
Former Yugoslavia	.0204	.0276	.0278	.8304	.0261	.0258	.8322
Poland	.0169	.0194	.0187	.2893	.0186	.0189	.9818
Romania	.0060	.0074	.0077	.7170	.0072	.0075	.06731
Former CIS	.0260	.0312	.0322	.2730	.0312	.0317	.7418
Arabic/Muslim states	.0122	.0177	.0181	.5674	.0170	.0175	.6667
Rest of the World	.0646	.0816	.0837	.3518	.0797	.0802	.08925
Fractionalization Index	.3547	.4461	.4519	.4174	.4335	.4371	.7511

3.6 Random Walk Algorithm

In the next step, specific anchor persons in households were drawn from each of the 250 PSUs sampled at the first stage. We aimed for a total of 2,500 interviewed migrant households. To ensure a sufficiently large gross sample, the number of migrant households to be drawn from each of the 250 PSUs was set to 80. Hence, 20,000 addresses (250 × 80) were to be drawn in total.

A distance-based and entirely simulated random walk procedure was implemented in the sampling process. This algorithm is based on a disproportional sampling scheme, which assigned higher sampling probabilities to selected migrant groups. In the following, details on the sampling scheme and on the random walk algorithm are being presented.

Household Selection and Household IDs

IEB-data at individual level was used in this step of the sampling procedure. As mentioned above, the enriched database contained information on PSU-membership as well as the geographical location, i.e. address coordinates.[12] As the SOEP is a household panel survey, the entire household of the sampled or selected individual is asked to take part in an interview. Therefore, for the sampling procedure, sampled individuals simultaneously represent their (so far unknown) households members.[13] In order to minimize the sampling of multiple household members, individuals were assigned to so-called "pseudo-households". These pseudo-households were created to identify individuals that were likely to be living in the same household. Two or more individuals were assigned to an identical pseudo-household ID if they shared the same address and surname. Once a specific pseudo-household ID was sampled, the random-walk algorithm did not sample another pseudo-household member, since the household has been already represented in the sample.

Disproportional Address Sampling

Disproportional address sampling was implemented to ensure specific migrant groups (both regarding country of origin and migrant generation) are sufficiently represented. The algorithm was based on a disproportional sampling scheme which assigned higher sampling probabilities to selected migrant groups. This approach guaranteed that a pre-defined minimum sample size for each of the targeted migrant groups was achieved, although some

[12]The geographic coordinates were linearly transformed so that a re-identification of addresses is virtually impossible.

[13]Hence, a distinction between individual and household level is not necessary to accomplish the target net sample size of 2,500 households.

of the groups were only poorly represented in the underlying population.

Table 5: Address Sampling Factors by Country of Origin

Country of Origin	Generation	Sampling Factor
Italy	1	8.67
	2	2.89
	n.d.	4.89
Spain/Greece	1	10.22
	2	6.22
	n.d.	6.44
Turkey	1	2.67
	2	1.11
	n.d.	1.56
Former Yugoslavia	1	3.33
	2	2.44
	n.d.	2.89
Late Repatriates	all	1.78*
Poland	all	3.33
Rumania	all	8.44
Former CIS	all	2.89*
Arabic Countries	all	1.56
Rest of World	all	1.00

*Please note that the sampling of addresses was slightly modified during the fieldwork period. As late repatriates and individuals from the former CIS showed substantially higher response rates, they were excluded from the second tranche at the end of the data collection. For more information see section 5.1 on design weighting.

Data from the German Microcensus 2009 (FSO 2010) was used to estimate the current population proportion across migrant groups. Afterwards a sampling factor was calculated to increase or decrease the sampling probability of specific groups (see Table 5). The procedure results in higher or lower sampling probabilities compared to the probabilities that would be obtained by a simple random sampling design (SRS).

Before sampling, the individual data-records in a given PSU were multiplied by their respective sampling probability. Sampling of anchor persons of households was based on these disproportional sampling factors ("sampling without replacement").

Starting Point

We randomly sampled 250 individuals (one person per PSU) with their corresponding geographic coordinates as starting points for our random walk procedure.[14] The random walk algorithm was separately processed in every single PSU.

Contrary to standard random walking approaches (see e.g., Hoffmeyer-Zlotnik 1997), the

[14]The random process was based on the "Mersenne-Twister" (see Matsumoto/Nishimura 1998).

applied random walk procedure was completely simulated.[15] For this purpose an algorithm was used to select addresses from a database containing the geocoded addresses of all members of the 250 sampled PSUs. The algorithm is based on a distance based sampling routine and an altering sampling interval.

Distance Based Sampling

Once a starting point was selected, address coordinates were used to calculate distances between the sampled starting point and all other individual addresses in a given PSU. The distances were used in order to sort the addresses prior to a varying interval sampling approach. The algorithm uses a linear distance to sample addresses. More specifically, "great circle distance" – based on the simple euclidean distance – is used to sort addresses depending on the distance from a sampled (starting) address.

The random walk algorithm constitutes a systematic sampling procedure based on a sorted-by-distance list of records. Hence, a sampling interval k was to be chosen so that every k^{th} element in the frame is selected. In our study, the initial sampling interval was set to 8. Therefore, at least 640 (80×8) addresses per PSU were needed to realize a sample of 80 households per PSU. The majority of PSUs met this condition. The selection procedure is repeated until the algorithm converges, i.e. because there is no next household left to be sampled. The number of selected households in a given PSU is counted and if the sum is below 80 – which is the intended number of households per sample point – the sampling procedure starts all over again using a smaller sampling interval ($k - 1$).

Figure 3 displays a visual representation of the address sampling in a given PSU. In this example, the PSU contains the geocoded locations of 2,210 individuals represented by dots. The randomly selected starting point is marked with a blue plus-symbol. The colored dots represent the addresses sampled by the random walk algorithm. As can be seen from the overview map on the lower left, the sampled households are located within a comparatively narrow spatial area. Thus, in the given example, the sampling algorithm succeeded in creating a random sample of addresses that is very-well suited for the use in the field since an interviewer has to travel comparatively short distances between sampled addresses.

[15] In a "standard" random walk (e.g. in the ADM sampling design (Hoffmeyer-Zlotnik 1997)) addresses are selected randomly using a street data base from which a starting address is randomly drawn. Following a specific scheme including rules on where to go when facing end-of-streets or highways, interviewers then select households, for instance by listing every fifth housing.

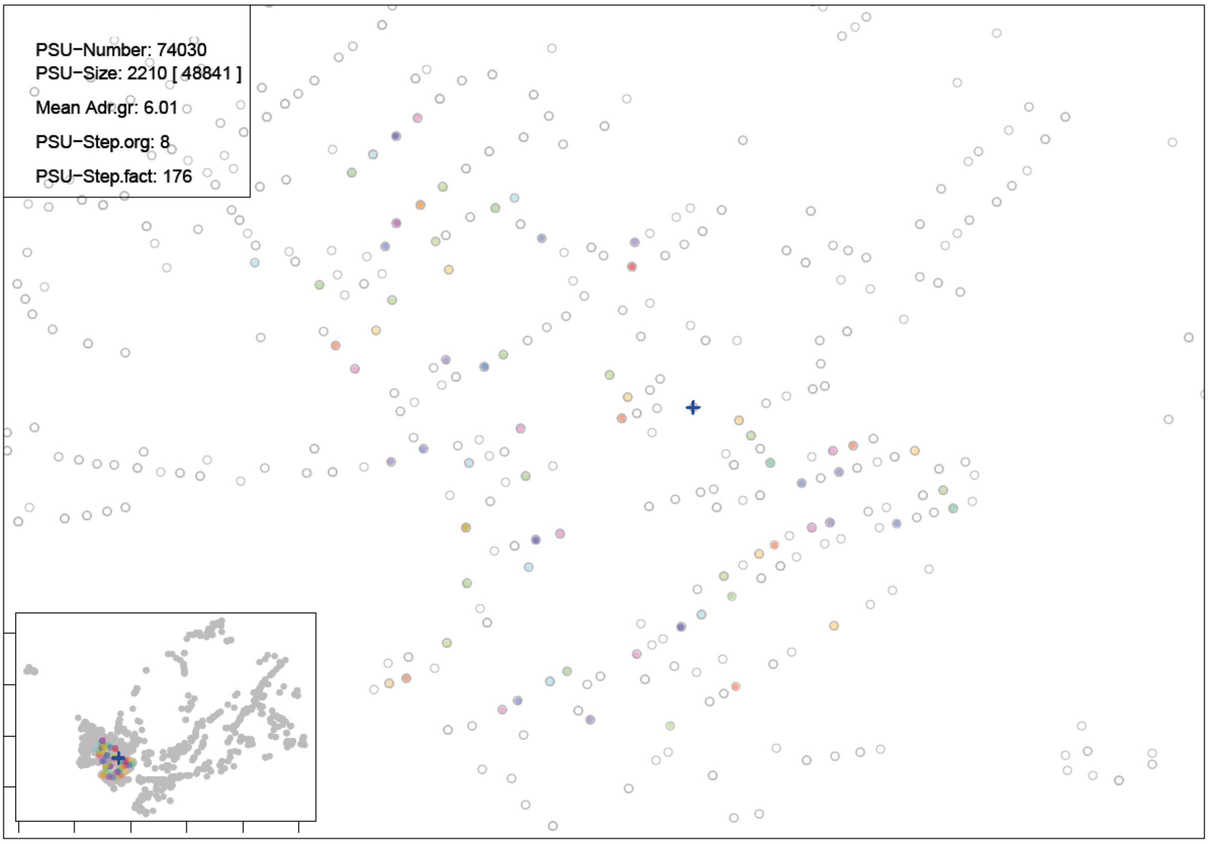

Figure 3: Visual Representation of Address Sampling in a Given PSU

Use of Sampled Households in Fieldwork

The 80 sampled addresses per PSU were split into different "sample tranches". The first tranche included 32 addresses and was given to the respective interviewer who was asked to interview 12 households. A second tranche included 16 addresses and was only used if the 12 interviews could have not been conducted by using the 32 addresses of the first tranche. The 32 remaining addresses functioned as a back-up for "quality-neutral nonresponse"[16].

3.7 Characteristics of the Gross Sample

The IAB-SOEP migration sample was intended to cover two groups: Those who immigrated since 1995 and those that are descendants of immigrants (second-generation migrants). As described above, a disproportional sampling design was implemented in order to control the composition of the sample in terms of countries of origin as well as migrant generation.

Table 6 displays the composition of the gross sample of 20,000 households (column 3) and the composition of the IEB population (column 2). Due to the disproportional sampling

[16]For example in cases where sampled individuals have already died or moved abroad.

probabilities, selected migrant groups meet higher (or lower) proportions compared to their representation in the sampling frame. For instance, the gross sample consists of 6.5% of migrants from Romania, although Romanian immigrants since 1995 merely represent the 2.8% of the IEB database.[17]

Table 6: Composition of Population and Samples by Nationality/Country of Origin

Nationality/Country of Origin	IEB since 1995		Gross Sample	
	n	%	n	%
Italy	168,733	4.6	1,764	8.8
Spain/Greece	119,365	3.3	1,892	9.5
Turkey	860,442	23.5	3,038	15.2
Former Yugoslavia	351,338	9.6	1,828	9.1
Late Repatriates	-	-	2,123	10.6
Poland	290,891	7.9	2,108	10.5
Romania	104,190	2.8	1,309	6.5
Former CIS	448,140	12.2	1,977	9.9
Arabic & Muslim	209,667	5.7	1,238	6.2
Rest of World	1,115,076	30.4	2,723	13.6
Subtotal	3,666,852	100.0	-	-
Germany	13,479,090	-	-	-
Total sample	17,415,942	-	20,000	100.0

[17] Please note that even though a sampling procedure is based on a randomized sampling of primary and secondary sampling units, the sampling procedure may nevertheless cause unintended systematic differences between the actual, empirical sampling probabilities. For instance, recent research on selection errors in (standard) random route samples (Bauer 2014) indicates that unequal empirical selection probabilities seem to be a general problem of random walk approaches. In the analysis of the influence of common random route instructions on empirical selection probabilities of addresses, the author revealed that some addresses are geographically located at more favourable positions – for instance in a city center – while others are located at less favourable positions, for instance, at a city border.

4 Results from Fieldwork and Response Rates

The IAB-SOEP Migration Sample (M1) was implemented in field from May to November of 2013. Target sample households were informed by a written announcement prior to the actual interview. The data collection was entirely based on computer-assisted personal interviews (CAPI). The interviews were conducted by a total of 217 interviewers who interviewed between 1 and 52 households with a mean of 12.5 households per interviewer.

Table 7: Final Results from Fieldwork (Households)

	n	%	n	%
Quality Neutral Drop-Out				
Invalid Address	342	2.6	-	-
Deceased	38	0.2	-	-
Moved Abroad	416	3.2	-	-
Subtotal	*796*	*6.1*	-	-
Response				
Full/Partial	2,723	20.9	2,723	22.3
Screenout	1,145	8.8	1,145	9.4
Subtotal	*3,868*	*29.8*	*3,868*	*31.7*
Nonresponse				
Refusal	4,584	35.3	4,584	37.6
Address Inquiry Unsuccessf.	2,037	15.7	2037	16.7
Non-contact	1,250	9.6	1,250	10.2
Language Problems	208	1.6	208	1.7
Other Reason	249	2.0	247	2.0
Subtotal	*8,328*	*64.1*	*8,328*	*68.3*
Total	12,992	100.0	12,196	100.0

Table 7 reports the final results of the fieldwork. The gross sample of 20,000 was processed in the fieldwork in distinct tranches. Not all of these tranches were used in the end, resulting in a reduced gross sample of 12,992 out of the 20,000 households sampled by the random walk procedure.[18] From the 12,992 households, 796 were classified as "quality neutral nonresponse" either because addresses were clearly invalid (342), individuals/households had been deceased (38), or they moved abroad (416).[19] From the reduced gross sample of 12,196 households, 3,868 households participated in a screening interview. Hence, the overall mean response rates of Sample M1 amounts to 31,7%.[20]

[18]The final field tranche of 1.000 households (anchors) omitted anchors classified as "late repatriate" and members of former CIS-states to compensate for the higher response rate in both groups (see, Table 8).

[19]The number of households moved abroad is fairly high (3.2%) compared to past SOEP samples. It seems reasonable to assume that this is largely due to re-migration processes.

[20]AAPOR Non-Response Definition RR2, see AAPOR (2011).

4.1 Response Rates by Regions and Country of Origin

Figure 4 displays response rates by federal states (left) and counties (right). As can bee seen from the maps, response rates displays notable regional variation. For instance, according to the map on the left, response rates seem to be higher in Eastern Germany. At county-level, variation in response rates is even greater, however, showing no distinct geographical pattern. These differences may in part also reflect the performance of interviewers allocated to these specific counties.

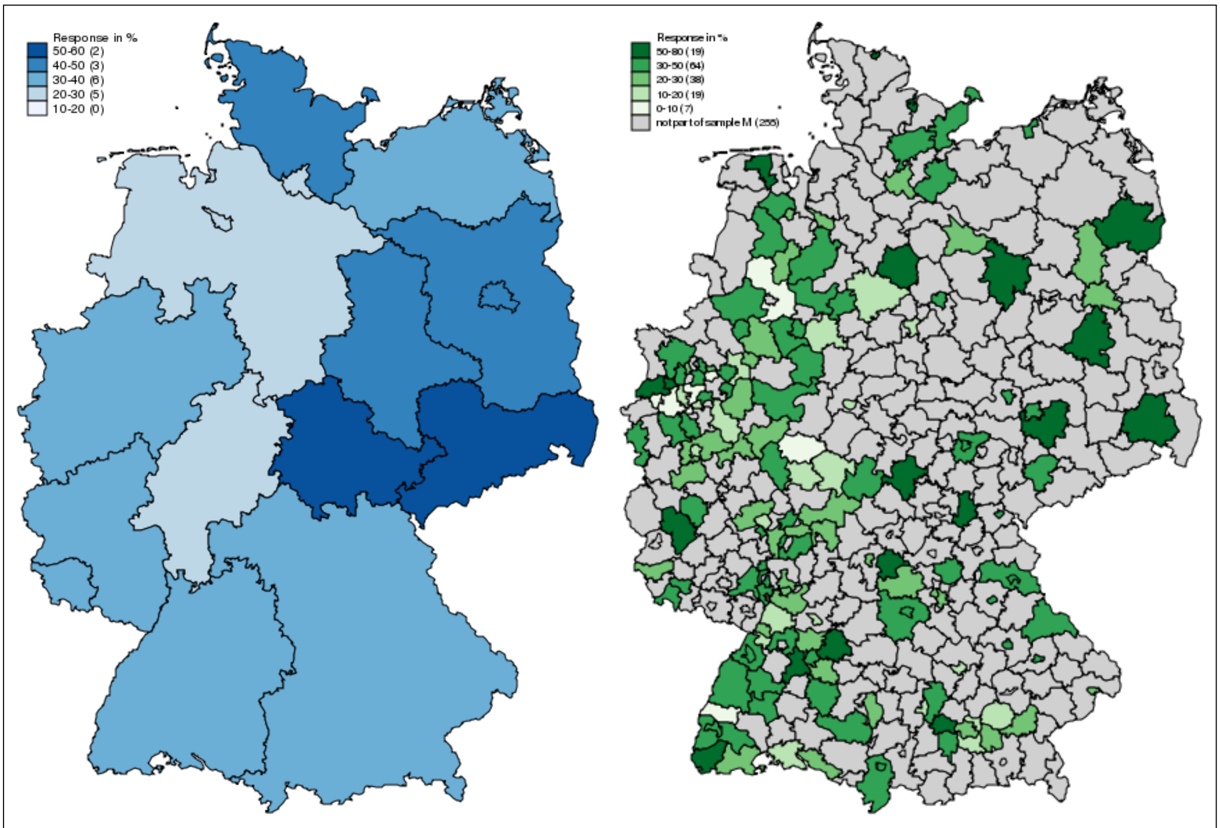

Figure 4: Response Rates by *Länder* and Counties

Furthermore, response rates also vary across migrant groups, as can be seen from Table 8. Subgroup response rates range from 25.7% in the group of migrants from Greece to 35.7% for migrants from former CIS-states. Also, the group of Late Repatriates yield a comparatively high response rate with 32.2% on average. Explaining the variation of response rates is one main task of the nonresponse weighting adjustment discussed in section 5.2.

Table 8: Response Rates by Migrant Groups

	Gross n (red.)	Net n	Response Rate (%)
Italy	1,082	314	29.0
Spain	394	121	30.7
Greece	794	204	25.7
Turkey	1,930	592	30.7
Former Yugoslavia	1,187	401	32.7
Late Repatriate	1,234	397	32.2
Poland	1,235	402	32.6
Rumania	759	230	30.3
Former CIS	1,150	422	35.7
Arabic Countries	803	226	28.1
Rest of World	1,628	559	34.3
Total	12,196	3,868	31.7

4.2 Screening of Target Population Members

The processing of addresses so far was based on our preliminary estimate of the persons being part of our sampling frame. The first set of questions in the interview validated these estimates. Interviews were only continued if the anchor person passed this screening. Screening was based on a few questions about the respondents' birth place, their year of immigration (if immigrated) as well as parental birthplace. Figure 5 displays the applied screening scheme.

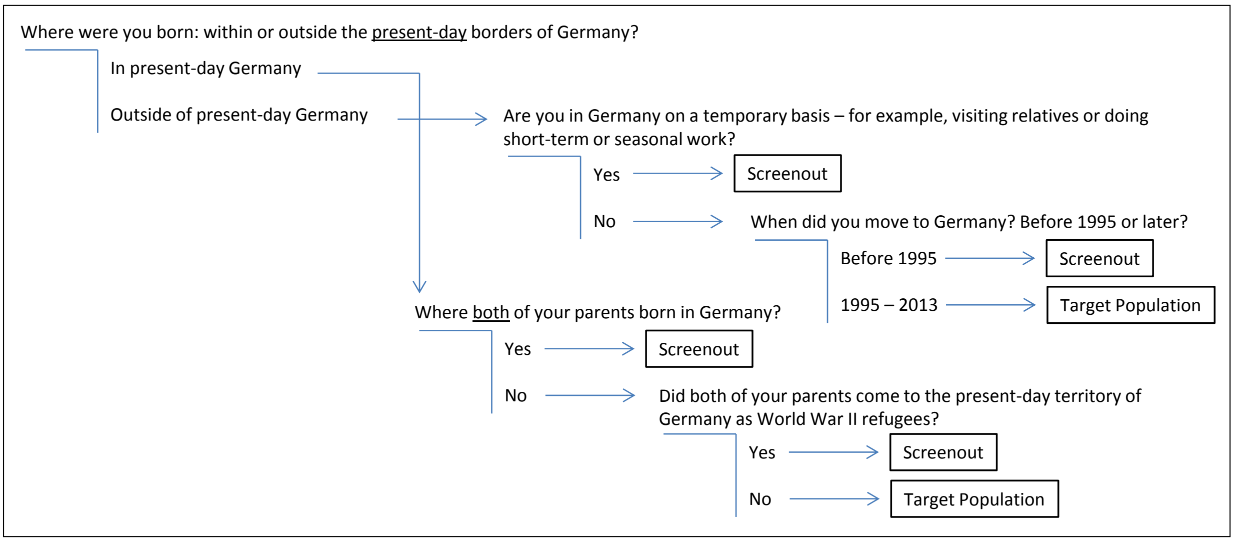

Figure 5: Screening and Identification of Target Population Anchors

Two target population sub-groups were identified during the screening interview: (1)

individuals born outside of Germany who immigrated into Germany between 1995 and 2013 and who were not in Germany only on a temporary basis and (2) individuals born in Germany with at least one parental unit born outside of Germany who did not came to Germany as a World War II refugee. Individuals who did not meet this definition were screened out, i.e. no survey interview was conducted and all other household members were disregarded as well.

Table 9 displays frequencies and total percentages for the different reasons for screenout grouped by country of origin. Altogether, 1,145 sampled households were screened out during fieldwork as it turned out they didn't make part of the target population either because they were seasonal laborers, they immigrated before 1995, their parents were both born in Germany or their parents immigrated into Germany as World War II refugees.

As can be seen from the totals, the main reason for screenout was immigration into Germany before 1995. As expected, some of the information on first IEB data entry used as a proxy for the individuals' immigration date was not valid. For instance, some immigrants probably did not immediately start working (as employees) and did not apply for benefits of the Federal Employment Agency. The second main reason for screenout concerns those anchors who did not immigrate into Germany and whose parents were both born in Germany as well. It seems reasonable to argue that these screenout cases are mainly due to a misspecification of migrant generation: Some of the sampled individuals classified as second-generation migrants are third-generation migrants or do not have a migration background at all. In addition, 57 anchors were screened out either because they were seasonal laborers or their parents immigrated into Germany as World War II refugees.

Table 9: Household Screenouts by Migrant Groups

Country of origin	Reasons for Screenout: Seasonal Laborer	Reasons for Screenout: Migration Before 1995	Reasons for Screenout: Parents: born in Germany	Reasons for Screenout: Parents: displaced persons	Total Screenout		Reduced Net Sample	
	n	n	n	n	n	%	n	%
Italy	3	42	66	1	112	9.8	202	7.4
Spain/Greece	2	52	42	0	96	8.4	229	3.1
Turkey	5	176	29	1	211	18.4	381	5.3
Former Yugosl.	3	106	27	3	139	10.7	262	9.6
Late Repatriate	2	29	19	0	50	4.4	347	12.7
Poland	16	73	43	1	133	11.6	269	9.9
Rumania	9	23	9	0	41	3.6	189	6.9
Former CIS	3	50	2	0	55	4.8	367	13.5
Arabic Countries	0	45	14	0	59	5.2	167	6.1
Rest of the World	3	62	179	5	249	21.7	310	11.4
Total	46	658	430	11	1,145	100.0	2,723	100.0

The 1,145 screened out households are not of interest in the following analysis of nonresponse and are therefore not taken into further consideration. All in all, the final, reduced gross sample for the nonresponse analysis consists of 11,051 households of which a total of 2,732 households was interviewed.

5 Cross-sectional Weighting of Sample M1

Like in most social science surveys, members of the target population do not have constant probability to be interviewed. One the one hand, this is due to decisions made by the researcher, for example by relying on a complex sampling design that assigns different selection probabilities to groups of the target population (*selection by design*). On the other hand, those unequal probabilities result from nonresponse of a subsample of individuals, e.g. those who refuse to take part in the survey (*self-selection*).

There are different strategies in order to deal with selective samples. Besides specialized model-based strategies (such as the Heckman selection model, see Heckman 1979) and the imputation of data (that is the replacing of missing data), the (ex-post) weighting of survey data is the most common way to handle selective samples to date. Different weighting procedures and techniques have been proposed over the last decades (Kalton/Flores-Cervantes 2003). One of the most common methods is referred to as "propensity score weighting" (Rosenbaum/Rubin 1983). Propensity score weighting approaches assign sample elements (such as households) with a greater "importance" if they hold characteristics associated with lower sample probabilities and lower response probabilities and vice versa. In order to do so, sample elements are weighted by the combination of their inverse sampling probability and their inverse response probability (conditional to being sampled in the first place). The combination allows for an unbiased estimation of population parameters. An example is the mean estimator developed by Horvitz/Thompson 1952:

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{i=1}^{N} \frac{s_i}{\pi_i \cdot P(x_i h \in S)} \tag{10}$$

In equation 5, π_i denotes the response probability of household i and $P(x_i h \in S)$ refers to the sampling probability of household i in stratum h. The variable s_i denotes a binary indicator, taking the value 1 if household i was observed, and 0 if household i was not observed.

In the following sections, we distinguish three steps in obtaining a cross-sectional weight of Sample M1. *Design-based weighting* corrects for unequal probabilities of sampling implemented by the researcher. In the next stage, (model-based) *nonresponse weighting adjustment* corrects for unequal response probabilities. These response probabilities are unknown and have to be estimated. We do so, first, by comparisons of the realized net sample with the margins of the gross sample and, second, by comparisons of the realized net sample with known margins of the underlying target population. The latter is usually referred to as *post-stratification* and *raking*.

5.1 Design Weighting

Design weights account for unequal sampling probabilities of households within sample points. As Section 3.5 shows, the two-stage sampling implements equal selection probabilities. Only in the subsequent step of the randomly selecting 80 persons per sample point, we introduced unequal sampling factors by country of origin and in/direct migration background (Section 3.6).

$$
w^{(1)}_{\text{design}}
\begin{cases}
8.67 & \text{if Origin} = \text{Italy and Generation} = 1 \\
2.89 & \text{if Origin} = \text{Italy and Generation} = 2 \\
4.89 & \text{if Origin} = \text{Italy and Generation} = \text{unclear} \\
10.22 & \text{if Origin} = \text{Spain/Greece and Generation} = 1 \\
6.22 & \text{if Origin} = \text{Spain/Greece and Generation} = 2 \\
6.44 & \text{if Origin} = \text{Spain/Greece and Generation} = \text{unclear} \\
2.67 & \text{if Origin} = \text{Turkey and Generation} = 1 \\
1.11 & \text{if Origin} = \text{Turkey and Generation} = 2 \\
1.56 & \text{if Origin} = \text{Turkey and Generation} = \text{unclear} \\
3.33 & \text{if Origin} = \text{Former Yugoslavia and Generation} = 1 \\
2.44 & \text{if Origin} = \text{Former Yugoslavia and Generation} = 2 \\
2.89 & \text{if Origin} = \text{Former Yugoslavia and Generation} = \text{unclear} \\
1.78 & \text{if Origin} = \text{Late Repatriate} \\
3.33 & \text{if Origin} = \text{Poland} \\
8.44 & \text{if Origin} = \text{Rumania} \\
2.89 & \text{if Origin} = \text{Former CIS} \\
1.56 & \text{if Origin} = \text{Arabic Countries} \\
1.00 & \text{if Origin} = \text{Rest of World}
\end{cases}
$$

Please note that SOEP is a household panel survey and the sampling frame of the IEB lists persons, not households. Hence, household with, for instance, two members of our target population have twice the sampling probability as a household that comprises only a single person from the underlying population. In a second step, we thus correct design weights by the number of persons in the household that match the definition of the target population[21] by multiplying $w^{(1)}_{\text{design}}$ with the inverse of the number of target population members per household n_{target} (see equation 11). Correction factors for households with more than three target population members were truncated at the bottom (max = 3) in order to limit the variance in weights:[22]

$$w^{(2)}_{\text{design}} \begin{cases} w^{(1)}_{\text{design}} \times \frac{1}{n_{target}}, & \text{if } n_{target} <= 3 \\ w^{(1)}_{\text{design}} \times \frac{1}{3}, & \text{if } n_{target} > 3 \end{cases} \tag{11}$$

With a final scalar multiplication, we construct design weights so that their sum over all units of the reduced gross sample of 11,051 equals the number of households in the target population estimated on the basis of the German Microcensus, namely 4.2 Mio. The final household design weights function as a basis for further weighting procedures. Characteristics and distributions of the weights are presented in section 6.

$$w^{(3)}_{\text{design}} = w^{(2)}_{\text{design}} * \frac{4,245,831}{\sum_{n=1}^{11,051} w^{(2)}_{\text{design}}} \tag{12}$$

[21]Target population members among the persons joining households with the anchor person were identified by using various information obtained in the conducted interviews. More specifically, individuals were classified as target population members if they were associated with one of the following groups: (i) the sampled anchors, (ii) those non-anchors being born outside of Germany and immigrated only after 1995, who are not in Germany only on a temporary basis (e.g. for seasonal work), and have been unemployed, employed (full-time, part-time) or in apprenticeship between 1995 and 2011, while being between 15 and 65 during that period of time. Also, second-generation migrants (born in Germany but not both parents born in Germany and not both parents came to Germany as displaced persons), and unemployed, employed (full-time, part-time) or in apprenticeship between 1995 and 2011, while being between 15 and 65 during that period of time and born in 1977 or later.

[22]The actual maximum number of target population members per household was 5 (4 households). 13 households contained 4 target population members.

5.2 Nonresponse Weighting Adjustment

In addition to the selectivity due to complex sample designs, further selectivity in the observed data may be introduced by the choice of the selected unit of analysis to participate in the survey or not (*(self)selection into surveys*).

Several theoretical explanations for unit nonresponse have been proposed to account for the self-selection into surveys. Most of them aim for the modeling of the individual's decision making process whether to approve or refuse participation (e.g. Groves et al. 1992). Following the rational-choice approach (e.g. Coleman/Fararo 1992) this decision can be regarded as a result of a cognitive evaluation of perceived costs and benefits of participation. Unfortunately, due to the non-participation itself, very little is known about nonrespondents (Giraldo/Zuanna 2006: 296) and their individual evaluations of costs and benefits are usually not measurable. Therefore, survey researchers aim to identify other variables, coming from alternative data sources, which might influence the individual's perception of costs and benefits of participation. The selection of variables used in the nonresponse weighting adjustment of Sample M1 is based on findings from other studies analyzing nonresponse. Besides information on sampled households provided by the interviewer (e.g. Olson 2006, Keeter et al. 2006, Abraham et al. 2006), we rely on geocoded information on the regional context (neighborhood, municipality) of sampled households (e.g. Johnson et al. 2006). Additionally, we make us of the rich IEB data available both at individual and enterprise level.

Over the last decades, response rates in social science surveys in Germany have been constantly decreasing. To date, only few studies achieve response rates above 40% (Schnell 2012: 164). For Sample M1, the response rate – as the share of realized interviews in the reduced gross sample (see AAPOR 2011) – amounts to 31.7% and is therefore only marginally lower compared to the latest SOEP samples, e.g. Sample J (33%) or Sample K (35%) (see Kroh et al. 2014). This small decrease is most likely due to the specific target population of migrants. Past research has already revealed that migrant groups yield lower response rates compared to non-migrants (e.g. Bethlehem et al. 2011: 64, Deding et al. 2008). From a rational choice perspective it seems reasonable to assume that the average perceived costs of participation are higher in migrant groups. For example, difficulties in understanding the receiving country's language may result in higher nonresponse rates as field material is not always provided in different languages. Furthermore, interviewers language skills might be limited.[23] Additionally, a lack of detailed knowledge about local institutions (e.g., federal agencies) may lead to higher feelings of suspicion and lower valuations of interviewers' (and sponsors') reliability, which would as well result in higher nonresponse rates. Another possible explanation may be confounding effects that operate through various socio-demographic characteristics such as education and social status and those being associated with both migration and participation probabilities.

If respondents and nonrespondents systematically differ from each other regarding specific survey variables, estimates may be biased. From a statistical perspective, nonresponse bias is a function of the response rate as well as differences between respondents and nonrespondents. This notion is based on the assumption of a "real" or "true" population parameter, a mean for example, which consists of the mean value of participants $\bar{x}_r$ and non-participants $\bar{x}_n$. Depending on the prevalence of unit nonresponse and the amount of variation between those groups, estimation results in a certain bias b (Bethlehem et al. 2011: 42):

$$b_{\bar{x}} = (\bar{x}_{response} - \bar{x}_{nonresponse}) \cdot \frac{n_{nonresponse}}{n_{total}} \tag{13}$$

The size of a nonresponse bias increases with decreasing response rates and/or increasing differences between participants and non-participants.[24]

5.2.1 Modeling Nonresponse

Various statistical techniques addressing nonresponse adjustment have been proposed within the last decades. Their common aim is to correct for those differences between respondents and non-respondents that cause biased estimates of population parameters. Model-based weighting methods such as *propensity score adjustment* use observable differences between the selected gross sample and the realized net sample. For this purpose information on both respondents *and* nonrespondents is used to calculate so-called "nonresponse weights". Using these nonresponse weights, the net sample distributions for

[23]Therefore, additional field material might not always be used even though necessary.

[24]Due to the fact that bias depends upon both nonresponse rate and difference between groups, relatively small response rates do not automatically cause estimators to be biased: If respondents and nonrespondents do not systematically vary from each other (with regard to specific survey variables), a bias will be small or even not present at all. Hence, the response rate itself holds no information on the presence or amount of bias. Groves/Peytcheva (2008) provide a meta-analysis on the impact of nonresponse rates on nonresponse bias.

various variables are weighted to meet the gross sample distributions for these variables (Kalton/Flores-Cervantes 2003: 83). In other words, nonresponse weights attach more "importance" to observed units of analysis (e.g. households) if they hold characteristics that are associated with nonresponse. In many cases, weighting procedures are complemented using other techniques that focus on differences between marginal distributions of the net sample and known marginal distributions of the underlying target population (e.g. post-stratification, raking, GREG).[25]

For SOEP Sample M1 data, the unknown participation probabilities were estimated using logistic regression which and then transformed into propensity weights (Kim/Kim 2007: 501f). A dummy variable for participation was constructed and used as the dependent variable.

The calculated nonresponse weights can be combined with design weights in order to correct parameter estimation of an underlying target population, namely private migrant households in Germany in 2013.

5.2.2 Correlates of Nonresponse: Data Sources

One of the key advantages of register-based samples – such as the given sample of migrants coming from the IEB – is the availability of information on nonrespondents at individual level, and on individual's information concerning sampled adresses. In both cases, we are able to rely on individual characteristics for the modeling of nonresponse in the given IAB-SOEP Migration Sample (M1).[26] Moreover, we use available aggregated, geographical data to improve non-response analyses. For this purpose we consider information coming from the Federal Statistical Office on municipalities (Gemeindestatistik) as well as counties (INKAR). Finally, we make use of commercial data from on neighborhoods (Microm data)

[25]Regardless of which technique is used to correct for the self-selection into the survey, assumptions about the nature of nonresponse need to be made. Usually, scholars distinguish between three missingness mechanisms, i.e. three types of missing data: Missing Completely at Random (MCAR), Missing at Random (MAR) and Missing not at Random (MNAR) (see Rubin 1976). The majority of weighting techniques assume that data is at least missing at random (MAR), i.e. nonresponse on a variable is related to other (observed) variables but not to values of the variable itself. In other words, some subgroups of the target population (with specific characteristics) are more or less likely to participate in a survey than others. However, when those differences in observed characteristics are controlled for, no systematic difference between respondents and nonrespondents (within groups) remain (Schafer 1997).

[26]Please note that the focus of our analysis lies upon the consistent estimation of response propensities. Hence, we are not primarily interested in a theoretical interpretation of effects (Spieß 2010: 123). With this information, a distinction has to be made between variables available for individual households on the one side, and aggregated spatial data linked to the sampled addresses of households on the other one. Particularly in the latter case, causal interpretations of significant relationships are hampered by the problem of ecological fallacies (McGaw/Watson 1976: 134f). Therefore, caution is needed in interpretation.

as well as IEB data aggregated at the level of sample points as well as enterprises in which employed respondents work.

Addresses: Field Work Information While field work for the address sampling process was carried out, SOEP interviewers collected information on the selected sampled households and their environment. This information provides useful insight about the living standards of sampled households. Residents of more expensive individual houses (possibly house owners), for example, are generally thought to be more prone to participate in a survey than people living in flats (Durrant/Steele 2009: 376), as studies have shown that wealthier individuals present higher participation probabilities (i.e. see. Abraham et al. 2006: 693f). For the modeling of nonresponse, we made use of information on the type of house resp. building (e.g., flat in multi-story building vs. individual house) and data on business intensity at the district level. Also, the size of the community (number of inhabitants) was coded[27] by the data collection enterprise (TNS Infratest).

Individual: IEB The use of data coming from the IEB for the ellaboration of the sampling frame comes along with major advantages for modeling nonresponse. Given the richnesss of the data, we are able to use detailed information at individual level for both respondents *and* nonrespondents. Beyond migration related information, the IEB provides socio-demographic and socio-economic variables such as age, gender or employment status that can be used to improve the modeling of nonresponse.

Enterprises: IEB Additionally, data from the IEB registers provides useful information on firms where the employed sample members are currently working.[28] We use information on the enterprises staff composition in terms of number of employees, their nationality, and job type.

PSU-level: IEB Furthermore, we draw on aggregated information on residential segregation of immigrants at PSU-level[29] as well as information on PSUs already used during the sampling of households such as the share of different migrant groups within

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

[28]Data from the "Establishment History Panel" (BHP) is linked to individuals during the merging process of building the IEB.

[29]The term segregation refers to the composition of characteristics hold by individuals in spatial areas (Boal 1976).

PSUs.

Neighborhood: Microm Data In addition to the information collected during fieldwork on the neighborhood characteristics of sampled households, we make use of information provided by the private enterprise "microm GmbH". The Microm data is linked to the SOEP data and may be used for guests and staff of the SOEP (see Goebel et al. 2007). This data provides detailed information on the local and regional characteristics concerning the social structure and environment/neighborhoods of households in Germany. The information is available at various levels of aggregation that range from the household-cell-level (in which a few households are grouped together), to that of the market-cells level (aprox. 470 households per cell) and then to 8-digit postal code districts (aprox. 500 households per district). Microm-Data contains detailed regional data for the analysis of nonresponse considering the social structures of households (e.g. family structure, education status) as well as the economic situation of households (e.g. purchasing power).

Municipality: Federal Statistical Office The "Regionaldatenbank Deutschland" (regional database Germany) contains information on different levels of geographic units that has been collected as part of a joint project of the Federal Statistics Office and its subnational counterparts at *Länder*-level. We make use of variables compiled at county and community level for the analysis of nonresponse. This set of variables concerns with data from the 2009 general parliamentary election (turnout, share for different parties), age structure and dwelling types.[30].

County: Federal Statistical Office (INKAR) The Federal Institute for Research on Building, Urban Affairs, and Spatial Development (BBSR) in Germany provides additional information on regional characteristics. The database "Indikatoren und Karten zur Raum- und Stadtentwicklung in Deutschland und in Europa" (INKAR) provides information on regional economic activity (e.g. prices for building grounds, household income, welfare benefits) as well as population characteristics (e.g. educational data) of the different regional entities. From this source, the information is available at the county-level and for NUTS-2[31] regions and compiled in 2009[32].

[30]For further information, see the web link under REGIONAL (2012) in the bibliography.

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

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

Additional Data Sources Finally, we make use of an additional variable on quota of party members at the county-level that was obtained from the "Deutscher Lernatlas" (Lernatlas 2011), a comparative research project on learning conditions and quality at regional county level.[33]

[33]Data can be downloaded freely without registration. See Lernatlas (2011) and `http://www.deutscher-lernatlas.de/ergebnisse/ueberblick/` [visited the 04th June 2015].

Table 10: List of Variables Used in Analysis of Nonresponse of Sample M1 – I

Variable	Source	Type	Values/ Range	Level	Year
Type of House	Field Information	Ordinal (4 Steps)	1= Individual 4= High Multi-story	Household	2013
Business Intensity (District)	Field Information	Ordinal (5 Steps)	1= Residential District 5=Industrial Zone	Household	2013
Municipality Size	Field Information	Ordinal (6 Steps)	1= <2k inh. 7= > 500k inh.	Municipality	2013
Business Intensity (Street)	Microm	Ordinal (6 Steps)	1= Accommodation Only 6 = Business Only	Street	2013
Household Structure	Microm	Ordinal (9 Steps)	1= Mainly Single Persons 9= Mainly families with children	House Cells	2013
Children per Household	Microm	Ordinal (9 steps)	1= Lowest value 9= Highest value 6= Average	House cells	2013
Status (Socio-economic)	Microm	Ordinal (9 Steps)	1= Loewest Status 9= Highest Status 5= Average	House Cells	2013
Share of College Graduates	Microm	Ordinal (7 Steps)	1= Below 2% 7= Above 35%	Street	2013
Exclusive Housing Environment	Microm	Binary	1= Yes 0=No	House Cells	2013
Unemployment	Microm	Ordinal (7 Steps)	1= Lowest 7= Highest 4= National Average	Postal Codes	2013
Prices for Building Grounds	Inkar	Metric	in €$/m^2$	County	2013
Average household income (per Person)	Inkar	Metric	in €	County	2013
GDP/capita	Inkar	Metric	in 1000 €'s	County	2013
Welfare Benefits for Renting Expenses	Inkar	Metric	in €	County	2013
Share of high School graduates	Inkar	Metric	-	County	2013
Electoral Turnout in 2009 General Election	Statistical Office	Metric	-	Municipality	2009
Vote Share for SPD	Statistical Office	Metric	-	Municipality	2009
Vote Share for CDU/CSU	Statistical Office	Metric	-	Municipality	2009
Vote Share for FDP	Statistical Office	Metric	-	Municipality	2009
Vote Share for *Alliance '90/ The Greens*	Statistical Office	Metric	-	Municipality	2009
Vote Share for *The Left*	Statistical Office	Metric	-	Municipality	2009
Vote Share for small parties	Statistical Office	Metric	-	Municipality	2009
Share of Small Flats (1-2 rooms)	Statistical Office	Metric	-	Municipality	2011
Share of Big Flats (6+ Rooms)	Statistical Office	Metric	-	Municipality	2011
Share of pop. aged 18-25	Statistical Office	Metric	-	Municipality	2011
Share of pop. aged 25-35	Statistical Office	Metric	-	Municipality	2011
Share of pop. aged 35-45	Statistical Office	Metric	-	Municipality	2011
Share of pop. aged 45-55	Statistical Office	Metric	-	Municipality	2011
Share of pop. aged 55-65	Statistical Office	Metric	-	Municipality	2011
Share of elderly (65+)	Statistical Office	Metric	-	Municipality	2011
Quota of Party Members	Lernatlas	Metric	-	County	2009

Table 11: List of Variables Used in Analysis of Nonresponse of Sample M1 – II

Variable	Source	Type	Values/ Range	level	Year
Segregation	IEB	Metric	-	PSU	2012
Share of Migrants in PSU	IEB	Metric	-	PSU	2012
Share of Migrant Groups in PSU	IEB	Set of metric variables	Turkish, Italian, Spanish, Greek, (...) etc.	PSU	2012
Country of Origin	IEB/ Onomastic	Set of Dummys	Turkish, Italian, Spanish, Greek, (...)	Household	2012
Migrant Generation	IEB	Set of dummies	1, 2, non differentiable	Individual	2012
Year of Birth	IEB	Metric	-	Individual	2012
Gender	IEB	Dummy	Female, Male	Individual	2012
Migration Status IEB	IEB	Metric	-	individual	2012
Late Repatriate IEB	IEB	Yes / No	-	individual	2012
Asylum Seeker	IEB	Yes / No	-	Individual	2012
Participation in German anguage Course	IEB	Yes / No	-	Individual	2012
Reception of Settling-in Grants	IEB	Yes / No	-	Individual	2012
IEB Data Source	IEB	Ordinal	Social Security Notifications (1), FEA Process Data (2), Both (3)	Individual	2012
Employment Status	IEB	Ordinal	Employed w/o. Welfare Benefits (1), Unemployed w. Welfare benefits (2), Employed w. Welfare Benefits (3), Unemployed w/o. Welfare Benefits, e.g. FEA Measure (4)	Individual	2012
Job Type	IEB	Dummy	Standard Employment vs. Rest	Individual	2012
Wage	IEB	Dummy	-	Individual	2012
Length of spell: Non-working Population	IEB	Metric	-	Individual	2012
Seniority	BHP	Dummy	<1 Year vs. >1 Year	Individual	2012
Number of Employees	BHP	Metric	-	Enterprise	2012
Number of Full-time Employees	BHP	Metric	-	Enterprise	2012
Number of Female Employees	BHP	Metric	-	Enterprise	2012
Median Wage	BHP	Metric	-	Enterprise	2012
Number of Low-qualified Employees	BHP	Metric	-	Enterprise	2012
Number of High-qualified Employees	BHP	Metric	-	Enterprise	2012
Number of Non-skilled Workers	BHP	Metric	-	Enterprise	2012
Number of Employees in Simple Jobs	BHP	Metric	-	Enterprise	2012
Number of Employees in Simple Services	BHP	Metric	-	Enterprise	2012
Number of Employees in Commercial or Administrative Jobs	BHP	Metric	-	Enterprise	2012
Sector / Industry	BHP	Ordinal	Energy and Mining (1), Agriculture (2), Manufacturing (3), Construction (4), Trade (5), Transport and Communications (6), Financial Intermediation (7), Other Services (8), Non-profits and Households (9), Public Administration (10)	Enterprise	2012

5.2.3 Multiple Imputation and Data Coding

For analysis of nonresponse and the generation of weights, it is necessary to have complete information for all observations included in the sample. Whenever this is not the case, the observations with missing information will be omitted from regression and weights will not be estimated for those observations.

Some of the selected variables contained missing values. In the majority of cases, all the values for all variables coming from one source were missing for a given spatial unit (county, municipality). However, none of the households yielded complete missings. In other words, in this case missing values do not cluster for one particular set of households. Furthermore, the overall share of missing information was considerably low, as can be seen from Table 12 in which we report the prevalence of missing data by groups of indicators and *Länder*.

Table 12: Share of Missings in Sample M1 for Different Variables by *Länder*

Variables	Field inf. (Address)	Microm (House-Cells)	Microm (Others)	INKAR (County)	Statistical Office (Municipality)	Partisanship (County)	IEB (Individual)	IEB/BHP (Enterprise)
Schleswig-Hol.	.0280	.0040	.0280	.0000	.0000	.0000	.0000	.0440
Hamburg	.0342	.0023	.0114	.0000	.0023	1.000	.0000	.0410
Lower Saxony	.0632	.0023	.0463	.0000	.0000	.0000	.0000	.0508
Bremen	.0214	.0071	.0071	.0000	.0071	.0000	.0000	.0571
Northrhine Westph.	.0256	.0020	.0205	.0000	.0000	.0266	.0000	.0570
Hesse	.0103	.0017	.0111	.0000	.0017	.0000	.0000	.1216
Rheinland-Palatinate	.0513	.0020	.0256	.0000	.0020	.0000	.0000	.0256
Baden-WuerT.	.0321	.0043	.0185	.0000	.0012	.0000	.0000	.0414
Bavaria	.0258	.0052	.0298	.0000	.0029	.0000	.0000	.0240
Saarland	.0394	.0000	.0079	.0000	.0000	.0000	.0000	.0787
Berlin	.0115	.0000	.0014	.0000	.0000	1.000	.0000	.0575
Brandenburg	.0227	.0227	.0455	.0000	.0000	.0000	.0000	.2424
Mecklenb.-Vorp.	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0222
Saxony	.0485	.0000	.0121	.0000	.0000	.0000	.0000	.1152
Saxony-Anhalt	.0000	.0079	.0394	.0000	.0000	.0000	.0000	.0394
Thuringia	.0274	.0137	.0411	.0000	.0137	.0000	.0000	.0274
Total	.0276	.0047	.0216	.0000	.0019	.1267	.0000	.0653

Missing observations were imputed to the dataset by way of the "Multiple imputation by chained equations" methodology which was implemented using the software *Stata* and its additional "ice"-program (Royston 2009). To account for the imputation uncertainty that this procedure implies, ten different predictions for the missing values were calculated (White et al. 2011: 378). Furthermore, the whole statistical procedure was implemented ten times with different starting values (Horton/Lipsitz 2001: 248). As a result, ten different complete datasets were available for analysis taking the uncertainty of multiple imputation into account via appropriate statistical routines (White et al. 2011: 377).

After the imputation procedure, some of the variables were transformed for the following analysis. Continuous variables were categorized, resulting in three distinct categories for most of the variables; in general, the middle category served as a reference group in the regression analysis. 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. Non-linear effects are controlled for, because for each group individual parameters are estimated. Also, this 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). And finally, interpretation and comparison of coefficients is more straightforward this way.

5.2.4 Results and Calculation of Nonresponse Weights

In order to model the households' nonresponse propensities, logit regressions were performed for different combinations of covariates using statistical routines that accounted for imputation uncertainty. Additionally, we used "robust standard errors" in order to account for possible heteroscedasticity and non-independent observations within sample points (see White 1980, Spieß 2010). A sample point identifier variable was used as a cluster variable in each of the estimated models.[34]

The full net sample of 2,723 households were used in every model. The initial full model included all available covariates displayed in Tables 10 and 11. As described above, most of the raw variables were categorized or transformed into dummies before inclusion into the full model. The second reduced model was estimated using only those variables that exert a significant effect (α = 5% level). Figure 6 displays the coefficients with their respective 95% confidence intervals for the reduced model. Only a small fraction of the initial covariates used in the full model reaches statistical significance. Hence, the reduced model is a lot more parsimonious than the full model. This is furthermore supported by the results for different criteria of model fit displayed in Table 13, showing no substantial differences between the full model and the reduced model.[35] Even though the full model

[34]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.

[35]Due to the lack of independence between observations in multiple imputed datasets, likelihood-based measures of model fit cannot be calculated appropriately. Therefore, the values for the reported *MacFadden's Pseudo-R^2* were calculated using a standard logistic regression approach with cluster-robust standard errors. Nonetheless, due to the small fraction of missing data, estimated parameters differed

fares slightly better compared to the reduced model, both models do not display good fit values. The results are however comparable to previous SOEP nonresponse analyses (e.g. Kroh et al. 2014).

Table 13: Fit Values for Estimated Models

	Full model	Reduced model
Pseudo-R^2	.0532	.0345
Error Rate	.2434	.2464

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

Overall, these results can be understood in a positive manner, especially regarding quality of sampling: A wide variety of variables from various data sources has been tested for their influence on response propensities, and only a small fraction of them has reached statistical significance. These results suggest that participation across groups indeed seems to be determined to a great extent by chance and that in many aspects, respondents and nonrespondents possibly do not systematically differ from each other.

However, turning to the individual coefficients that are displayed in figure 6, some substantial differences remain and require further consideration.

Migration: The majority of covariates capturing the migration background do not hold significant effects. However, anchor persons classified as late repatriates and migrants from former CIS-states show higher participation rates compared to other migrant groups. Contrary to what was expected, second-generation migrants seem to be less reluctant to participate in the survey when compared to respondents who were classified as first-generation immigrants or those for whom it was not possible to classify definitely. As expected, the year of the first IEB data entry – used as a proxy for the year of immigration – is positively related to participation propensities: More recent immigrated households were more likely to take part of the survey. In conclusion, the results regarding migration related variables can be understood in a positive manner: A great number of variables have been tested for their influence on participation and only few of them yield significant (usually small) effects. Hence, participation propensities seem to be quite comparable between different migrant subgroups.

only marginally.

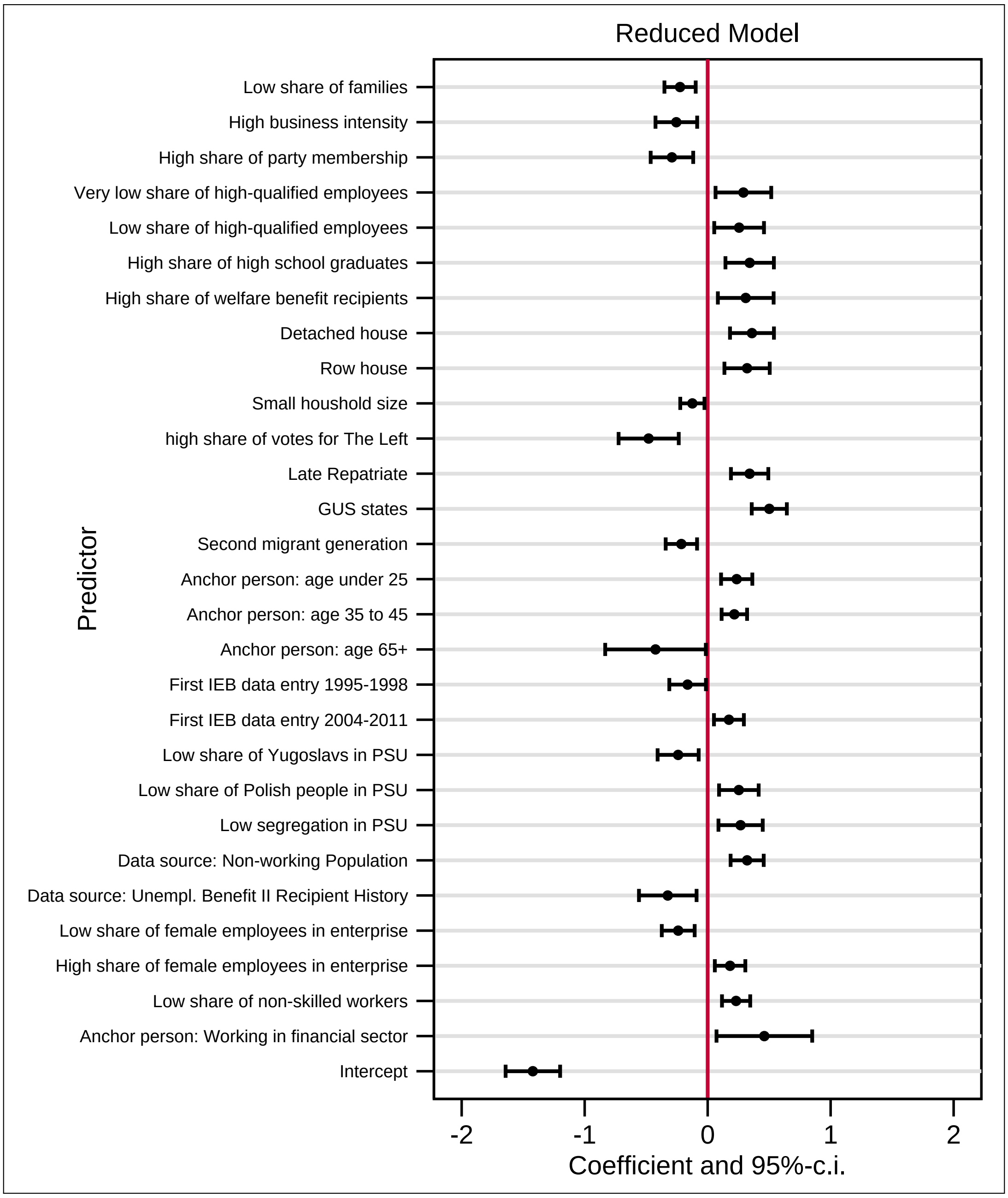

Figure 6: Coefficients and Confidence Intervals for the Estimated Reduced Model

Note: The dependent variable was coded 1 for participation and 0 for non-participation. Number of observations in the model n = 11,051

Education: As expected, the educational level of a region's inhabitants is related to participation probabilities. Response probabilities seem to be higher in areas with a high share of high school graduates. This is in line with results from other studies (e.g. Abraham et al. 2006, Durrant/Steele 2009) and previous SOEP analysis of nonresponse (Kroh et al. 2014).

Economic Conditions: The overall picture for coefficients relating to the economic conditions of households (in an area) indicates higher response probabilities in economical weaker areas. Areas with a high share of welfare benefit recipients hold significantly higher response rates. The observed relationship is furthermore supported by coefficients at individual level: Anchor persons sampled from the IEB non-working population hold significantly higher response rates.

Local/Regional Characteristics: Response rates are significantly lower in areas characterized by a high business intensity. As expected, households residing in detached houses or row-houses are more likely to participate. Furthermore, households in areas with low segregation display significantly higher response rates. In addition, response rates are significantly lower in sampling points with a low share of Yugoslavs while being significantly higher in sampling points with a low share of Polish people.

Socio-demographics: The respondents' age seems to be related to response propensities. Individuals aged under 25 as well as individuals aged 35 to 45 are more reluctant to participate in the survey. In contrast, anchor persons aged 65+ display significantly lower response rates. These results are in line with other studies (Keeter et al. 2006). The respondents' gender yielded no significant effect.

Enterprise: The majority of enterprise-level characteristics of employed respondents yielded no significant effects. However, respondents employed in enterprises with a high share of female employees showed higher response propensities. Furthermore, a low share of non-skilled workers in a respondent's enterprise is associated with higher response rates. In addition, respondents working in enterprises within the financial sector showed higher response rates. Please note that the mechanisms causing the observed relationships are probably complex and hard to explain. The obtained effects are likely to be (partially) influenced by other unobserved variables. Nevertheless, it becomes evident that controlling for multiple enterprise- and work-related variables is reasonable, especially in work- and employment-related surveys.

Further Variables: Contrary to our expectations, the share of party membership in an area is negatively related to participation probabilities. However, comparable effects were observed in the nonresponse adjustment of Sample K data (Kroh et al. 2014) in which areas with a high election turnout showed lower participation rates as well. Furthermore, a high share of election votes for the party *Die Linke (The Left)* in an area yielded a negative effect on participation probabilities.

The reduced model (as displayed in figure 6) was used to predict household participation probabilities.[36]

Table 14 displays a comparison between actual response rates and mean estimated response probabilities by sample point. Whereas the model fit values did not indicate strong predictive power of the considered variables, the correlation between the two variables ($\rho_{(obs;est)} = 0.48$) reveals that the 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.

Table 14: Comparison of Estimated and Actual Response Rates by Sample Point

Response Rates	Min.	*p25*	Median	*p75*	Max.
Observed (x_{obs})	00.00	15.69	23.86	33.33	64.00
Estimated (x_{est})	12.91	21.11	24.74	28.60	48.34

Note: At the sample point level the two variables correlate with $\rho_{(obs;est)} = 0.48$

The nonresponse weights for further analysis are calculated as inverse response probabilities.

Table 15: Characteristics of Raw Estimated Nonresponse Weights

Min.	*p10*	*p25*	Median	*p75*	*p90*	Max.	Mean	SD
1.47	2.77	3.36	4.25	5.42	6.85	16.63	4.59	1.73

Note: All values were rounded to the second decimal place.

Table 15 displays characteristics of the raw estimated nonresponse weights. Estimated weights, even after having been reduced in variation through regression design, still cover a wide range and rather extreme differences are obtained between the value of the 90%-percentile (6.85) and the corresponding maximum value of the distribution (16.63), which

[36]The model appears suitable for analysis since its effects have proven to be robust among different regression specifications and it does not contain insignificant predictors artificially increasing the variation of estimated probabilities and thereby also nonresponse weights.

signals the presence of outliers in the data. As mentioned before, large variation in weights can be harmful as variation of estimates may increase substantially (van Goor/Stuiver 1998). The trimming of weights (e.g., Peytchev et al. 2011) to be more equally distributed seems reasonable in order to counteract possible undesired features of applying weights in estimation applications. However, trimming may result in a small loss of efficiency. There are various possible trimming-approaches and the selection of a specific procedure is best based on the ratio of estimated weights.[37] For the given sample, we aimed household weights not to exceed 1.75 times the weight's mean:

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

The trimming procedure affected a total of 95 households with a weight above approx. 7.1. Hence, trimming succeeded as variation in weights was decreased while only a few weights had to be corrected and therefore holding loss of efficiency at a minimum level.

The estimated nonresponse weights can be tested by using the known data both on respondents and nonrespondents. In order to test whether weighting may reduce bias in estimation, unweighted and weighted mean estimates using only participating households (2,723) are compared to the actual value estimated using data for all households of the reduced gross sample (11,051). Weighting of the net sample was performed by using the product of design weights and nonresponse weights as a combined weight. The gross sample was weighted relying exclusively on design weights resulting in what is referred to as the "actual" values of the variables of interest. Table 16 displays the results for various variables used in estimation.[38] Bias was calculated as percent deviation from the actual value.

As can be seen, the weighted estimates' bias is smaller for all variables compared to the unweighted bias. In other words, weighted estimates (means) of the observed net sample are closer to the actual values observed in the gross sample. After being weighted, the net sample yielded no significant differences compared to the gross sample (actual values). Hence, estimation was improved using the estimated weights.[39]

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

[38] Calculations were performed with the first of the imputed datasets. Household-specific means of estimated weights were calculated and used in analysis. Trimming was implemented only after this procedure.

[39] An alternative strategy to test the performance of weights is their use in a regression application.

Table 16: Comparison of Weighted and Unweighted Estimates and Reduction of Bias

Variable	Actual value	Estimate´s Bias in Percent: Unweighted	Estimate´s Bias in Percent: Weighted	Reduction of Bias
Share of Pop. Aged 18-25	8.46	0.34	0.01	90 %
Share of Pop. Aged 65+	9.29	0.95***	0.02	75 %
Average Unemployment Benefit Expenses	787.42	0.43***	0.01	68 %
Average Household Income	1,632.73	0.36	0.03	91 %
Share of High School Graduates	31.66	0.91*	0.10	87 %
Prices for Building Grounds	244.17	2.68	1.74	35 %
Share of Party Membership	1.03	0.90	0.43	56 %
Share of Unemployed	6.29	1.05	0.92	13 %
Share of Turkish Anchor Persons	20.24	6.63***	1.06	84 %
Turnover in Accommodation	6.48	2.13***	0.13	94 %
Share of Single Households	41.27	1.19***	0.01	94 %
Share of Employees in the Tertiary Sector	70.03	1.38***	0.17	87 %

Note: *'s indicate result of two-tailed t-test for differences in means (difference from real value).
$* \rightarrow p < .1; \quad ** \rightarrow p < .5; \quad *** \rightarrow p < .01;$

Drawing upon the test proposed by DuMouchel/Duncan (1983), weights can be tested for their significant impact on regression parameters. The test was originally developed for the assessment of model specification and tests parameters in weighted and unweighted regression against each other. The test itself uses weights and interactions with the predictor. The resulting coefficients are then assessed in an F-test, in which significance indicates that weights cannot be ignored.

5.3 Poststratification and Raking

In addition to the reported nonresponse weighting adjustment, household weights were corrected using *post-stratification* and *raking* so that the Sample M1 meets known cell distributions resp. marginals. In standard post-stratification or cell weighting procedures, weights are adjusted so that given sample totals fit to known cell distributions from the underlying target population "on a cell-by-cell basis" (Kalton/Flores-Cervantes 2003). Raking – also referred to as "iterative proportional fitting" (Deming/Stephan 1940) – is a special case of post-stratification and is used "when poststrata are formed using more than one variable, but only the marginal population totals are known" (Lohr 2010: 344).

The post-stratification procedure completes the three step process of calculating so-called "first wave weights" for new SOEP samples, thus constituting a combination of design weights, nonresponse weights and post-stratification procedures. First wave weights are available for all SOEP subsamples A to M1. 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.

Final Sample M1 first wave household weights are stored in the variable `BDHHRFM`.[40] First wave weights at individual (person) level are stored in the variable `BDPHRFM` respectively.[41]

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 2013. The Microcensus is conducted by the Federal Statistical Office of Germany (FSO) and is a one-percent sample of the German resident population. For the first time Microcensus benchmarks of 2013 are based on the results of the German Census in 2011 ("Zensus 2011").

Table 17 lists the characteristics at the household level that were used in the raking-process of Sample M1. In addition to the characteristics used also for the SOEP as a whole – household size, *Länder*, municipality size and house ownership –, Sample M1 weights were adjusted with respect to a number of migration related characteristics as well. More precisely, we combined information on migrant generation, year of immigration (in the case of first generation immigrants) and country of origin in order to classify both households and individuals.

[40] For more information on SOEP variables and datasets see Haisken-DeNew/Frick (2005).

[41] 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.

Table 17: Population Characteristics Used in the SOEP Raking Procedure at Household Level

Variable	Values
Migration Group of Anchor in Target	1. Gen, 1995-2004, Turkey 1. Gen, 1995-2004, Spain/Greece/Italy 1. Gen, 1995-2004, Poland 1. Gen, 1995-2004, Former CIS 1. Gen, 1995-2004, Arabic 1. Gen, 1995-2004, Late Repatriate 1. Gen, 1995-2004, ROW 1. Gen, 2005+, Spain/Greece/Italy 1. Gen, 2005+, Poland 1. Gen, 2005+, Former CIS 1. Gen, 2005+, Rest of World 2. Gen, Not Turkey 2. Gen, Turkey
Household Size in Target	1 / 2 / 3 / 4 / 5+
Länder in Target	Baden-Württemberg / Bavaria / Berlin / Brandenburg / Bremen / Hamburg / Hesse / Lower Saxony / Mecklenburg-Vorpommern / North Rhine-Westphalia / Rhineland-Palatinate / Saarland / Saxony / Saxony-Anhalt / Schleswig-Holstein / Thuringia
Municipality Size in Target	< 20.000 / 20.000-100.000 / 100.000-500.000 / > 500.000
Home Ownership in Target	Yes / No

Note: Population characteristics were derived from the Microcensus 2013 (see FSO 2015).

The first step of the post-stratification procedure was an identification of target population households, both in the Sample M1 data as well as in the Microcensus 2013 database. Households were defined as target population members if they hold one of the two following conditions:[42]

1. at least one household member immigrated since 1995 or
2. at least one household member is a second-generation migrant and was born between 1977 and 1996.

[42] The restriction on age replicates the restriction implemented in the procedure of reducing the IEB database prior to onomastics analysis and sampling procedures (see section 3.1). Note that a comparison is subject to small limitations due to different definitions and concepts of "migration" used in the Microcensus and the IAB-SOEP Migration Sample (M1). Comparability between the two datasets was maximized wherever possible. However, some differences remain between the generated variables in two databases. For example, we made use of information on parents (past) citizenship as a proxy for country of birth if parental information on birthplace was not available in the Microcensus.

For the post-stratification at household level, sampled households needed to be clearly classified in terms of migrant generation, immigration year and country of origin. As households regularly consist of multiple target population members, decisions had to be made on how to unambiguously classify households. We made use of the concept of household representatives in order to classify households.[43] The household representative selection strategy was performed both for Sample M1 data and Microcensus data.

Table 18: Comparison of Strata Composition Prior to Post-Stratification (Household Representatives)

Strata	Microcensus 2013		Sample M1	
	n	%	n	%
1. Generation				
imm. 1995-2004, Turkey	181,546	4.28	154	5.66
imm. 1995-2004, Spain/Greece/Italy	62,045	1.46	142	5.21
imm. 1995-2004, Poland	99,518	2.34	128	4.70
imm. 1995-2004, Former CIS	309,573	7.29	292	10.72
imm. 1995-2004, Arabic Countries	94,758	2.23	76	2.79
imm. 1995-2004, Late Repatriate	396,667	9.34	338	12.41
imm. 1995-2004, Rest of World	625,472	14.73	376	13.81
imm. 2005+, Turkey/Spain/Greece/Italy	61,710	2.27	135	4.96
imm. 2005+, Poland	125,016	2.94	98	3.60
imm. 2005+, GUS	96,366	2.27	91	3.34
imm. 2005+, Rest of World	675,159	15.90	328	12.05
2. Generation				
Turkey	402,481	9.48	172	6.32
Rest of World	1,115,520	26.27	393	14.43
Total	4,245,831	100.00	2,723	100.00

Afterwards, households were categorized into strata. Those strata represent various combinations

and (3) country of origin. Table 18 displays unweighted frequencies and percentages over strata for both Sample M1 and Microcensus 2013 data. As expected, substantial differences are being observed for those migration groups that were selected with higher probabilities within the disproportional sampling design, for instance, immigrants from Spain, Greece or Italy.

[43]Household representatives pass their individual characteristics to the household as a whole. Decisive rules were set up in order to select an individual as a representative. First, IEB anchors were preferred over other household members. Second, immigrants were preferred over second-generation migrants. If there were multiple immigrants in a household, we chose the household member with the most recent year of immigration. In the case of duplicate immigration years, female household members were preferred over male members. On the next stage we considered the current employment status, preferring full-time employment over part-time employment, and part-time employment over unemployment. Finally, in the rare case of multiple potential household representatives at this stage of the selection procedure, a household member was randomly selected.

Subsequent to the raking at household level, Sample M1 data was additionally post-stratified at person level in order to generate individual first-wave weights. Again, information on population totals was derived from the German Microcensus 2013. Raking at individual level is based on the combined first wave households weights (the result of the raking procedure at household level) and uses information on the SOEP standard post-stratification characteristics household type, gender and age (see table 19) as well as migration related characteristics such as the migrant group (which has been already used in the raking procedure at household level).

Table 19: Population Characteristics Used in the SOEP Raking Procedure at Individual Level

Variable	**Values**
Migration Group	1. Gen, < 1995, Turkey 1. Gen, < 1995, Spain/Greece/Italy 1. Gen, < 1995, Late Repatriate 1. Gen, < 1995, Rest of World 1. Gen, 1995-2004, Turkey 1. Gen, 1995-2004, Spain/Greece/Italy 1. Gen, 1995-2004, Poland 1. Gen, 1995-2004, Former CIS 1. Gen, 1995-2004, Arabic 1. Gen, 1995-2004, Late Repatriate 1. Gen, 1995-2004, ROW 1. Gen, 2005+, Spain/Greece/Italy 1. Gen, 2005+, Poland 1. Gen, 2005+, Former CIS 1. Gen, 2005+, Rest of World 2. Gen, Not Turkey 2. Gen, Turkey Native German
Migrant Generation / Year of inmigration	1. Gen, prior 1995 / 1. Gen, 1995-2004 / 1. Gen, 2005+ / 2. Gen / Native
Sex	Male / Female
Age	0-20 / 20-25 / 25-30 / 35-40 / 40-45 / 50-55 / 55-60 / 60-65 / 65+

Note: Population characteristics were derived from the Microcensus 2013 (see FSO 2015).

6 Characteristics of Combined Cross-Sectional Weights

As mentioned above, cross-sectional weighting of first-wave SOEP data is a three step process: design weighting (1), nonresponse adjustment (2) and post-stratification (3). A combination of all three steps results in so-called "first wave weights" which are available for all SOEP subsamples A to M1.

Table 20 displays characteristics of household weights in each of the three weighting steps of Sample M1. The initial design weights of Sample M1 were grossed up so that the number of households in the sample meets the total number of 4.2 Mio. households in the underlying target population (see section 5.1).

Table 20: Characteristics of Weights During the Weighting Process

		Quantiles								
	Min.	10%	25%	50%	75%	90%	Max.	Mean	SD	n
Design Weight	32	113	170	299	490	882	980	384	270	2,723
Design*Nonresponse	92	384	632	1,144	2,040	3,544	8,005	1,565	1,317	2,723
Combined First Wave Weight	66	331	560	1,032	1,995	3,498	9,173	1,564	1,529	2,723

Variance in design weights is due to the disproportional sampling of households in accordance to country of origin and migrant generation (see chapter 3.6). At the second stage, design weights were combined (multiplied) with nonresponse weights. As can be seen from the quantiles and standard deviations, variation in weights increases as expected. In a third step, the product of design and nonresponse weights was post-stratified, resulting in the final combined first wave weights for Sample M1. The raking procedure introduces additional variation into the weights, however, only to a limited extent.

The distribution of the weights at all three steps is also displayed in Figure 7. As can be seen, weights are fairly right-skewed, both prior to raking (orange line) and post raking (blue line).

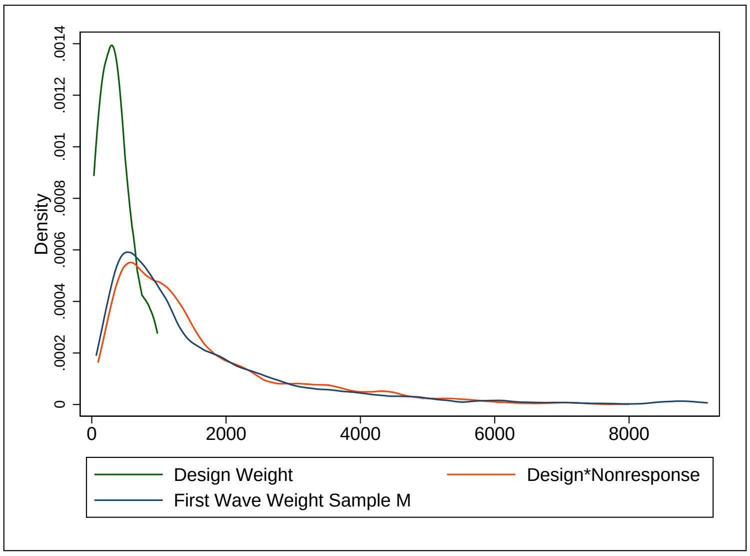

Figure 7: Distribution of Weights at all Three Weighting-Steps

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 household heads 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 households 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). Some researchers speak of a "middle class bias" (e.g., Goyder et al. 2002). 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 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).

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 cost of participation. Furthermore, past literature points to the fact that moving may make houesehold 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 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 highly educated individuals display 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. (2006: 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 M1 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 an important predictor in many social sciences research settings and controlling for it in weight generation seems reasonable.

IEB

Country of origin and migrant generation: The country of origin and migrant generation were formed on the basis of the IEB variables and an onomastic procedure. Generally, it has been shown that migrant groups differ in their responses rates (e.g., Babka von Gostomski/Pupeter 2008).

Year of first IEB data entry: We expected individuals who only recently immigrated to be less likely to participate because of two main reasons: First, recent immigrants may change their place of residence more frequently as it usually takes some time to find a more permanent place in German society. Second, recent immigrants may be more skeptical towards the survey topic and sponsor as they usually have less information on German institutions and public authorities.

Year of birth and gender: Past research revealed that older people are less likely to participate in a survey (Johnson et al. 2006). Furthermore, it has been shown that males are more likely to refuse participation (Smith 1983).

Household size: Single households were expected to be more likely to refuse compared to multiple-person households (Groves/Couper 1998).

Segregation in PSU: PSU segregation level may function as a proxy information for social integration. Those who feel less integrated may be furthermore less prone to participate in a social survey.

Share of migrants (and migrant groups) in PSU: Migrants tend to be, on average, of lower socio-economic status compared to natives. Hence, the share of migrants in a given PSU may function as an additional proxy of the area's wealth and attractiveness.

BHP

Wage: → See "Status (socio-economic)" [Microm-Data]

Seniority, Sector/industry, Staff Composition: We included information on sector/industry as well as enterprises' staff composition in the analysis of nonresponse. Even though we did not expect substantive effects at this aggregated level, an integration of those characteristics did seem reasonable as the IAB-SOEP Migration Sample M1 is thought to specifically enable researchers to analyze employment and employment biographies.

Further Variables

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.

References

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

Abraham, K. G./A. Maitland/S. M. Bianchi (2006). "Nonresponse in the American Time Use Survey". In: *Public Opinion Quarterly* 70(5), pp. 676–703.

BAMF (2013). *Migrationsbericht des Bundesamtes für Migration und Flüchtlinge im Auftrag der Bundesregierung. Migrationsbericht 2011. Berlin.*

Babka von Gostomski, C./M. Pupeter (2008). "Zufallsbefragung von Ausländern auf Basis des Ausländerzentralregisters". In: *Methoden – Daten – Analysen* 2(2), pp. 149–177.

Bauer, J. (2014). "Selection Errors of Random Route Samples". In: *Sociological Methods & Research* 43(3), pp. 519–544.

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

Boal, F. W. (1976). "Ethnic residential segregation". In: *Social areas in cities* 1, pp. 41–79.

Brücker, H./M. Kroh/S. Bartsch/J. Goebel/S. Kühne/E. Liebau/P. Trübswetter/I. Tucci/-J. Schupp (2014). *The New IAB-SOEP Migration Sample: An Introduction into the Methodology and the Contents. SOEP Survey Papers No. 216, Series C – Data Documentations.*

Coleman, J./T. Fararo (1992). *Rational choice theory*. New York: Sage.

Deding, M./T. Fridberg/V. Jakobsen (2008). "Non-response in a survey among immigrants in Denmark". In: *Survey Research Methods* 2(3), pp. 107–121.

Deming, W./F. Stephan (1940). "On a least squares Adjustment of a sample frequency table when the expected marginal total are known". In: *Journal of the American Statistical Association* 35, pp. 615–630.

Diehl, C. (2007). *Materialband und Endbericht zur Neuzuwandererbefragung - Pilotstudie: 1. und 2. Welle. Materialien zur Bevölkerungswissenschaft, No. 122. Wiesbaden: Bundesinstitut für Bevölkerungsforschung (BIB).*

DuMouchel, W. H./G. 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, G. B./F. 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.

Eisnecker, P./K. Erhardt/P. Trübswetter (2015). *The IAB-SOEP Migration Sample – Linking Survey to Administrative Data. Forthcoming.*

FSO (2010). *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2009. Fachserie 1 Reihe 2.2.*

— (2013). *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2012. Fachserie 1 Reihe 2.2.*

— (2015). *Bevölkerung und Erwerbstätigkeit. Vorläufige Ergebnisse der Bevölkerungsfortschreibung auf Grundlage des Zensus 2011.*

Gelman, A./J. Hill (2007). *Data analysis using regression and multilevel/hierarchical models*. Analytical methods for social research. Cambridge University Press.

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

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

Goyder, J./K. Warriner/S. Miller (2002). “Evaluation Socio-economic Status (SES) Bias in Survey Nonresponse”. In: *Journal of Official Statistics* 18(1), pp. 1–11.

Groves, R. M./M. P. Couper (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.

Groves, R. M./E. Peytcheva (2008). “The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis”. In: *Public Opinion Quarterly* 72(2), pp. 167–189.

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

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

Heckman, J. (1979). "Sample Selection Bias as a Specification Error". In: *Econometrica* 47(1), pp. 153–161.

Hoffmeyer-Zlotnik, J. H. (1997). "Random-Route-Stichproben nach ADM". In: *Stichproben in der Umfragepraxis*. Ed. by S. Gabler/J. H. Hoffmeyer-Zlotnik. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 33–42.

Horton, N. J./S. 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, A./K. Schneiderheinze (2013). *Linguistic analysis of personal names based on th "Onomastic process". www.stichproben.de/herunterladen/onomastic_process_HS_GbR.pdf Letzter Abruf: 20.02.2014.*

INKAR (2011). *Indikatoren und Karten zur Raum- und Stadtentwicklung in Deutschland und in Europa*. URL: http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/InteraktiveAnwendungen/INKAR/inkar_online_node.html (visited on 06/01/2015).

Jacobebbinghaus, P./S. Seth (2007). "The German Integrated Employment Biographies Sample IEBS". In: *Schmollers Jahrbuch* 127(2), pp. 335–342.

Johnson, T. P./Y. I. Cho/R. T. Campbell/A. 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, G./I. Flores-Cervantes (2003). "Weighting Methods". In: *Journal of Official Statistics* 19(2), pp. 81–97.

Keeter, S./C. Kennedy/M. Dimock/J. Best/P. 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, J. K./J. J. Kim (2007). “Nonresponse weighting adjustment using estimated response probability”. In: *Canadian Journal of Statistics* 35(4), pp. 501–514.

Kroh, M./K. Käppner/S. Kühne (2014). *Sampling, Nonresponse, and Weighting in the 2011 and 2012 Refreshment Samples J and K of the Socio-Economic Panel. SOEP Survey Papers No. 260, Series C – Data Documentations. DIW/SOEP, Berlin.*

Kroh, M./R. Siegers/S. Kühne (2015). “Gewichtung und Integration von Auffrischungsstichproben am Beispiel des Sozio-oekonomischen Panels (SOEP) (Weighting and Integration of Refreshment Samples: The Case of the SOEP)”. In: *Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen.* Ed. by J. Schupp/C. Wolf. Wiesbaden: Springer VS.

Lernatlas (2011). *Deutscher Lernatlas.* URL: `http://www.deutscher-lernatlas.de/ergebnisse/ueberblick/` (visited on 06/01/2015).

Liebau, E. (2011). “Arbeitsmarktintegration von hochqualifizierten Zuwanderern: Erklärung des spezifischen Integrationsmusters in den deutschen Arbeitsmarkt von Aussiedlern und jüdischen Kontingentflüchtlingen aus der ehemaligen Sowjetunion”. PhD thesis. University of Mannheim.

Lipps, O./K. Kissau (2012). “Nonresponse in an Individual Register Sample Telephone Survey in Lucerne/Switzerland”. In: *Telephone Surveys in Europe: Research and Practice.* Ed. by M. Häder/S. Häder/M. Kühne. Heidelberg: Springer.

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

Matsumoto, M./T. Nishimura (1998). “Mersenne twister. A 623-dimnensionally equidistributed uniform pseudorandom number generator”. In: *ACM Transactions on Modeling and Computer Simulation* 3, pp. 3–30.

McGaw, D./G. Watson (1976). *Political and social inquiry.* New York: Wiley.

Oberschachtsiek, D./P. Scioch/C. Seysen/J. Heining (2009). *Integrated Employment Biographies Sample (IEBS). Handbook for the IEBS in th 2008 version.* Nürnberg.

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

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

REGIONAL (2012). *Regionaldatenbank Deutschland.* URL: `https://www.regionalstatistik.de/genesis/online;jsessionid=3798563B386EC36254C9997A78F78D57?Menu=Impressum` (visited on 06/01/2015).

Rosenbaum, P./D. Rubin (1983). “The central role of the propensity score in observational studies for causal effects”. In: *Biometrika* 70, pp. 41–55.

Royston, P. (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.

Rubin, D. B. (1976). “Inference and Missing Data”. In: *Biometrika* 63(3), pp. 581–592.

Salentin, K. (1999). “Die Stichprobenziehung bei Zuwandererbefragungen”. In: *ZUMA-Nachrichten* 45(23), pp. 115–135.

Santacreu Fernandez, O/N. Rother/M. Braun (2006). “Stichprobenziehung für Migrantenpopulationen in fünf Ländern: eine Darstellung des methodischen Vorgehens im PIONEUR-Projekt”. In: *ZUMA-Nachrichten* 30(59), pp. 72–88.

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

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

— (2013). *Linking Surveys and Administrative Data. WP-GRLC-2013-03.*

Schnell, R./T. Gramlich/T. Bachteler/J. Reiher/M. Trappmann/M. Smid/I. Becher (2013). *A new Name-Based Sampling Method for Migrants unsing n-grams. WP-GRLC-2013-04.* Working Paper.

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

Smith, F. (1983). “On the Validity of Inference from Non-Random Samples”. In: *Journal of the Royal Statistical Society, Series A* 146, pp. 394–403.

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

TNS Infratest (2013). *SOEP 2012 – Methodenbericht zum Befragungsjahr 2012 (Welle 29) des Sozio-oekonomischen Panels. SOEP Survey Papers 144: Series B. Berlin: DIW/SOEP.*

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

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