Deliverable Number: 2.7

Deliverable Title: Scoping study of the auxiliary data sources and its potential to reduce nonresponse bias

Work Package: WP2 – Representing the population

Deliverable type: Report

Dissemination status: Public

Submitted by: SHARE ERIC

Authors: Senta-Melissa Pflüger (MEA), Michael Bergmann (MEA), Johanna Bristle (MEA)

Date Submitted: June, 2018

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654221.

www.seriss.eu @SERISS_EU

SERISS (Synergies for Europe's Research Infrastructures in the Social Sciences) aims to exploit synergies, foster collaboration and develop shared standards between Europe's social science infrastructures in order to better equip these infrastructures to play a major role in addressing Europe's grand societal challenges and ensure that European policymaking is built on a solid base of the highest-quality socio-economic evidence.

The four year project (2015-19) is a collaboration between the three leading European Research Infrastructures in the social sciences – the European Social Survey (ESS ERIC), the Survey of Health Ageing and Retirement in Europe (SHARE ERIC) and the Consortium of European Social Science Data Archives (CESSDA AS) – and organisations representing the Generations and Gender Programme (GGP), European Values Study (EVS) and the WageIndicator Survey.

Work focuses on three key areas: Addressing key challenges for cross-national data collection, breaking down barriers between social science infrastructures and embracing the future of the social sciences.

Please cite this deliverable as: Pflüger et al. (2018). *Scoping study of the auxiliary data sources and its potential to reduce nonresponse bias*. Deliverable 2.7 of the SERISS project funded under the *European Union's Horizon 2020 research and innovation programme* GA No: 654221. Available at: www.seriss.eu/resources/deliverables

Deliverable 2.7: Scoping study of the auxiliary data sources and its potential to reduce nonresponse bias

Senta-Melissa Pflüger, Michael Bergmann, and Johanna Bristle

Contents

Summary

Bias due to unit nonresponse is seen as one of the major error sources in survey data. Successful adjustment for this type of error is heavily dependent on the availability of suitable auxiliary variables. This report provides an overview of different types of auxiliary data and their potential for nonresponse bias adjustment. Therefore, in a first part, a scoping study with partly simulated data from Germany is conducted in which selected auxiliary variables are examined regarding their usability in nonresponse bias adjustment. Two different adjustment methods, response propensity weighting and the bivariate probit model with sample selection, are applied and evaluated regarding their potential for bias reduction. The second part of the report focuses on findings about different (cross-national) data sources of auxiliary variables. The indicators are discussed along the dimensions of content, accessibility, and quality. Our main findings are: the results of the scoping study show that auxiliary variables are useful in bias adjustment if they are (strongly) correlated with both the response indicator and the survey variable of interest. Whether variables fulfilling these requirements can be found in practice is uncertain and remains a primary challenge for future research. Concerning the availability of auxiliary variables, in principle many different sources exist, their measurement, quality, and accessibility, however, vary across countries and institutions which makes them less applicable at cross-national level.

1 Introduction

This report is part of SERISS work package 2 “Representing the population” which focuses on strategies before and after data collection to obtain high-quality samples in social surveys. The potential of such sample surveys to reflect human behavior in large populations strongly depends on the quality and completeness of the collected data. However, almost all surveys suffer from some amount of nonresponse, which is why it is seen as one of the most relevant error sources in survey data. Moreover, the topic is of increasing importance because of declining survey response rates over the last three decades (De Heer & De Leeuw, 2002; Groves & Couper, 1998; Steeh, 1981) and decreasing willingness of individuals to participate in surveys (Singer, 2006). The cross-national surveys within the SERISS project (European Social Survey (ESS), European Value Study (EVS), Generation and Gender Program (GGP), and the Survey of Health, Ageing and Retirement in Europe (SHARE)) are no exception to this situation. Rather they constitute specific challenges associated with using auxiliary data to study nonresponse bias. Deliverable D2.7 will, therefore, conduct a scoping study of the auxiliary data sources and its potential to reduce nonresponse bias. Based on substantial preparatory work as well as previous research in this area, this report takes SHARE as an example to study the potential of auxiliary variables for nonresponse bias adjustments. The results of this scoping study can then be used to inform other surveys within the SERISS project regarding the usefulness of available data sources for bias reduction.

The remainder of this report is structured as follows: The next chapter will give some general background information on nonresponse. Afterward, chapter 3 will present a simulation study partly based on SHARE wave 5 data regarding the usage of regional auxiliary variables in the adjustment of nonresponse bias in a specific estimate. Thereby, two different adjustment methods based on the used auxiliary variables are explained theoretically and applied to the estimation before the results are discussed. For a closer examination of the possibilities in nonresponse bias analysis, chapter 4 then discusses available data sources with regard to accessibility, quality, and potential use for nonresponse bias adjustment. The last chapter summarizes and evaluates the results of the report and points out possible alternatives especially for cross-national surveys.

2 Nonresponse in cross-national surveys

Nonresponse is one of the major components of the Total Survey Error (TSE) framework, which comprises several error sources in survey estimates (Groves et al., 2009). In principle, nonresponse can be split into two types, *unit nonresponse* and *item nonresponse*. Unit nonresponse describes the situation in which sampled individuals are not taking part in the interview, either because they were not contacted successfully or because they directly refused to be interviewed. While non-contact formed a large component of total nonresponse in earlier times, nowadays refusals constitute the major part of the total nonresponse in most surveys (Brick & Williams, 2013). In this situation, no survey information about the nonrespondents is available. In the case of item nonresponse, on the contrary, sampled units do participate in the survey, but do not provide a valid answer to all questions. This report focuses on missing data due to unit nonresponse.

Although nonresponse is an omnipresent phenomenon, missing observations are often neglected in the analysis of survey data. It is, however, important to think about the impact of nonresponse on the estimates derived from survey data as this influences the ability to generalize the results to the underlying population. When missing values are ignored and estimates are calculated based on the responding subsample only, two problems may emerge. The first is the loss of precision which is due to the reduced sample size and constitutes a threat for statistical inference. The second, and likely more problematic, issue is the bias which may occur in specific estimates derived from the survey data. If respondents and nonrespondents vary systematically with regard to the variable of interest, the responding units cannot be seen as a random subsample of the initially drawn sample and estimates will most likely be biased. The magnitude of the bias depends on two things: The difference between respondents and nonrespondents and the response rate.

Since only little information about nonrespondents is available in most cases, auxiliary variables are needed for nonresponse bias analyses. According to Groves (2006) and other researchers (Kalton & Maligalig, 1991; Little, 1986; Little & Vartivarian, 2003, 2005; Peytchev & Olson, 2007; Särndal & Lundström, 2005), auxiliary data should meet several criteria to be able to adjust for nonresponse bias. First, the values of all the additionally used variables should be non-missing for the entire sample, respondents and nonrespondents. Second, ideally, the data should be measured without measurement error (i.e. should be a reliable data source). Further requirements concern the association of the auxiliary variables with the constructs of interest. It is commonly stated that only variables which are correlated with the response indicator (R), or more general with the tendency to respond to a study, *and* the key variable of interest (Y) are useful in adjustments for nonresponse bias. If the variables are not or only weakly correlated, using them may not reduce nonresponse bias, but increase the variance of the estimate. Little & Vartivarian (2005) argue that the correlation coefficient of both associations should lie between $\rho = |0.48|$ and $\rho = |0.8|$ which describes quite high correlations. Whether auxiliary variables with such strong relations can be found in practice is questionable since it is, in general, difficult to find useful auxiliary information. Even in analyses of single countries, only relatively small relationships have been examined (Kreuter et al., 2010b; Kreuter & Olson, 2011; West, 2013; West & Little, 2013). However, Peytchev & Olson (2007) show that even small correlations can produce a change in the estimates, as long as the auxiliary variable is correlated with both the response indicator and the key variable of interest, which makes it worthwhile to search for auxiliary data even in a cross-national perspective.

In practice, approaches to limit nonresponse bias are conducted in three phases: before, during and after data collection. The strategies in the first area include, for example, careful training of the interviewers which can help to reduce nonresponse (Lessler & Kalsbeek, 1992). Procedures applied during fieldwork phase place an emphasis on obtaining high response rates. Sometimes specific features like increased incentives (Singer, 2002), responsive design (Groves & Heeringa, 2006; Schouten et al., 2013) or follow-up surveys (Glynn et al., 1993; Graham & Donaldson, 1993) are adopted. The third category relates to adjusting for nonresponse after the data have been collected using various adjustment strategies. In this report, the focus is set on the last field, the post-survey adjustments for unit nonresponse based on suitable auxiliary variables.

To explore auxiliary data sources in order to reduce nonresponse bias, we will first conduct a simulation study for one country only. Based on our findings, it might then be easier to properly evaluate if efforts in this direction should be emphasized and the search and use of auxiliary data have a positive cost-benefit ratio, also regarding cross-national studies.

3 Scoping study - evaluation of auxiliary data in nonresponse bias adjustment

3.1 Background and simulation setting

To evaluate the potential of one specific set of auxiliary variables in the reduction of nonresponse bias, this chapter presents a scoping study with partly simulated data. The simulated dataset is based on data from the "Survey of Health, Ageing and Retirement in Europe" (SHARE, Börsch-Supan et al., 2013)[1] which is a multi-disciplinary and international panel study targeting individuals aged 50 years and

[1] Acknowledgement: This paper uses data from SHARE Wave 5 (DOI: 10.6103/SHARE.w5.600) The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging ($U01\_AG09740-13S2, P01\_AG005842, P01\_AG08291, P30\_AG12815, R21\_AG025169, Y1-AG-4553-01, IAG\_BSR06-11, OGHA\_04-064, HHSN271201300071C$) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

older and their partners. For the purpose of this report, only data from the fifth wave of SHARE in Germany are used. These data can be combined with regional information from external data sources in Germany via a regional code included in both datasets.

The setting of this analysis is cross-sectional since the response process of the first wave substantially differs from responses in later waves (Lepkowski & Couper, 2002). The employed auxiliary indicators are derived from two sources, the German Census, conducted in 2011, and the “Indicators, Maps and Graphics on Spatial, and Urban Monitoring” (INKAR) collected from the Federal Institute for Research on Building, Urban Affairs, and Spatial Development.[2] The information is measured at the municipality or county level, which correspond to the LAU2 or NUTS3 regional code respectively (for an overview of regional codes see http://ec.europa.eu/eurostat/web/nuts/overview) and include different indicators like the population size in a certain municipality, the percentage of women/ men, the unemployment rate or the population density. A list of all selected variables can be found in the appendix (see Table A.1). The selection is based on considerations of their completeness (available for all respondents and nonrespondents in the sample) and their relation to response behavior and information gathered in SHARE. The Census data were collected in 2011 and the indicators from INKAR are taken from the years 2012 or 2013 to match the year of the data collection of Wave 5 of SHARE in 2012/2013.

In general, nonresponse bias is seen as a variable-specific error and not as universal for all variables within one study (Groves, 2006). Thus, specific exploration for each possible dependent variable of interest is required. We picked the variable “homeownership” as an illustrative example which showed significant correlations with the auxiliary variables in the SHARE data. Therefore, the analysis of this report is concentrated on the estimation of the proportion of the binary variable “homeowner”, which has the value 1 if the responding person (or her partner) owns the accommodation the respondents are living in. The reference category (i.e. 0) is comprised of the other categories “member of a cooperative”, “tenant”, “subtenant”, and “rent free”. After some restrictions, the original SHARE sample including the added auxiliary variables consists of 10616 individuals from which 4276 individual interviews were conducted in the years 2012 and 2013.

[2]data accessible for free on the referring homepages: https://ergebnisse.zensus2011.de/ and http://www.inkar.de/

One general problem in measuring the exact bias in the estimates due to nonresponse is the lack of reference values in the target population which leads to a restriction in the interpretation of the adjusted results since one cannot know whether the adjustments shift the estimates in the right direction. Therefore, simulation studies can help to evaluate the performance of various adjustment methods since the population value is known in this case. To preserve the possibility of linking the auxiliary variables via the regional code and build a dataset which is as realistic as possible, the SHARE data are used as a basis. The idea is to treat the respondent set (N = 4276) as a new reference population and use the referring estimates as true population values. Then, new samples of size n = 3000 are drawn from this population and unit nonresponse is introduced by manually deleting the survey information of some individuals (auxiliary variables are kept). The amount of nonrespondents is set to 60% to make it comparable to the initial SHARE data. This leads to a consistent sample composition of 1200 respondents and 1800 nonrespondents.

Most of the adjustment methods, especially weighting procedures, depend on the assumption of Missing at Random (MAR) data (Rubin, 1976). In this data situation, the response probability does not depend on the individual values of the key variable, but rather depends on some or all of the observed auxiliary variables. This means that there is a systematic relationship between the response propensity and the observed data and that the arising bias due to nonresponse can be adjusted for by including the auxiliary variables in the estimation. To introduce MAR missing observations to the data at hand, information from the initial SHARE dataset, including respondents and nonrespondents, about the missing behavior is used to model a missing data pattern which is as realistic as possible. To allow for relationships between incorporated variables, logistic regressions are applied to model the probability of responding. The dependent variable is an indicator which is 0 if the sampled individual is a nonrespondent and 1 otherwise. Auxiliary variables, chosen according to their importance in explaining response behavior and the variable of interest, are used as independent factors. Only those variables which are at least moderately correlated ($\rho > |0.1|$) to the response indicator *and* the variable of interest are included in the model (see Figure B.1 and B.2 in the appendix for an overview of the correlations of all auxiliary variables in the original SHARE data). In principle, it would suffice to use variables which are predictive for survey response. Since the aim of the analysis is to evaluate the potential of certain variables in nonresponse bias reduction and it has been shown that correlation with the variable of interest is important for reducing bias, this choice of covariates is applied. To build the missing at random mechanism, individual response propensities are predicted from the estimation results and divided into two groups at the 75%-quartile of the distribution (due to the large number of induced nonrespondents). Afterward, the 1800 nonrespondents are chosen randomly from the individuals in the group with a low predicted response probability and the survey information of these individuals is deleted. This leads to a situation in which the newly built response indicator is dependent on the covariates used in the logistic regression model of the SHARE data and, hence, detailed knowledge about the response process is available.

To provide a general overview of the distribution of the interesting estimate, T = 100 random samples of size n = 3000 are drawn by simple random sampling without replacement from the newly built target population of N = 4276 individuals. The reported estimates of the proportion in the results chapter (see chapter 3.2.2.2) are then always averaged over these 100 samples.

In the next chapter, the potential of the examined auxiliary variables is evaluated in more detail. As a first step of analysis, the correlations between the constructs in the newly built datasets are described before specific adjustment methods are presented and applied.

3.2 The potential of auxiliary variables in nonresponse bias adjustment

3.2.1 Basic criteria

As mentioned earlier, useful auxiliary variables for adjusting nonresponse bias should be correlated with both, the response indicator and the survey variable of interest. In the following, the associations in the simulated datasets are analyzed. The parameters refer to the situation of MAR missing data with 60% missing observations and the auxiliary variables, which are included in the introduction of the MAR missing data and are used in the adjustment methods described below. The results for the correlations between the response indicator (R) and the “homeowner” variable (Y) are shown in Table 3.1 below. The correlation coefficient displayed is the average over the estimates from the 100 iterations.

Auxiliary variable	Correlation with R	Correlation with Y
Municipality size	-0.4104	-0.2003
% houses inhabited by owner	0.3956	0.2063
% household type: married couple	0.4309	0.2162
% households with senior and younger citizens	0.5191	0.1342

Table 3.1: Average correlations between the auxiliary variables and the response indicator (R) and the "homeowner" variable (Y) in the 100 simulated datasets

In general, the average correlations of the response indicator with each of the auxiliary variables are quite high and also meet the criteria assessed by Little & Vartivarian (2005). The signs of the correlations are identical to those of the correlations in the original SHARE data (see Figure B.1 and Figure B.2 in the appendix). According to the requirements for auxiliary data presented earlier, the indicators in this data situation should be able to reduce nonresponse bias in the adjustments. It should, however, be kept in mind that the comparatively high correlations observed are a product of the missing data introduction and are rarely achieved in practice due to the lack of "good" auxiliary variables. In the next chapter, two adjustment methods based on the variables described here are proposed and applied to the simulated data to evaluate their potential for nonresponse bias reduction.

3.2.2 Adjustment methods based on auxiliary data

According to Little & Rubin (2002) the adjustments for nonresponse bias can generally be categorized into four groups: ad-hoc methods, weighting adjustment, imputation, and maximum-likelihood based methods. In this paper, the focus is on one weighting procedure, response propensity weighting (Little, 1986), and one maximum-likelihood based method, the bivariate probit model with sample selection (Van de Ven & Van Praag, 1981) which can be seen as an extension of the Heckman Selection Model (Heckman, 1976, 1979). The ad-hoc methods are used as a reference in the comparisons because they represent the situation in which nonresponse is not taken into account.

To evaluate whether auxiliary variables are useful in the adjustment for unit nonresponse bias, these adjustment methods are applied to the estimation of the proportion of homeowners and the results are compared regarding the following indicators. As a first step, the absolute bias is observed, which is calculated as the deviation of the sample estimate from the population value, which is known in the simulation study. According to the results from Collins et al. (2001) in evaluating the performance of different adjustment methods, it is mandatory to consider other indicators aside from bias reduction since effects may differ across criteria. It has, for example, been shown that most of the adjustment procedures (especially weighting methods) lead to an increased variance of the estimates (Kish, 1965, 1992). As a result, confidence intervals are wider and the ability to detect significant results is reduced. Little & Vartivarian (2005) state, however, that weighting can also decrease the variance of an estimate if the adjustment variables are correlated with the survey outcome of interest. Therefore, measures of variation of the estimates should also be evaluated and the trade-off between reduced bias and increased variance should be analyzed. Ideally, adjustments shift estimates in the direction of the true value, hence reduce bias, while maintaining or even reducing the estimate's variance. To combine both factors, the change in bias as well as in variance, the Mean Squared Error (MSE) can be used as a measure of the accuracy of the estimate (Wolter, 1985). The lower the MSE is, the smaller impacts of error can be expected. Let $\hat{\theta}$ be the estimate of the parameter of interest. Then, the MSE is defined by the following equation combining the systematic (bias) and the random error term (variance): $MSE(\hat{\theta}) = Bias(\hat{\theta})^2 + Var(\hat{\theta})$.

In the next chapter, the two selected methods are explained theoretically before applying them to the simulated data and presenting the results.

3.2.2.1 Theory and application

Weighting approaches

Weighting adjustments are frequently used when compensating for unit nonresponse. The idea is to weight each responding individual with a specific factor so that it also represents the sampled individuals not taking part in the survey. These methods are generally based on the assumption of MAR missing data, which was explained earlier.

The basis for the estimation of a proportion in the weighting approaches is the commonly used Horvitz-Thompson estimator (Horvitz & Thompson, 1952) for the population mean

$$\hat{\bar{Y}}^{HT} = \frac{\sum_{i=1}^{n_R} w_i y_i}{\sum_{i=1}^{n_R} w_i},$$

where n_R denotes the number of responding individuals and w_i describes the individual weighting factor which includes adjustments for the sampling design and/or nonresponse. Due to the non-linear form of the estimator of interest, the variance estimation for the weighted estimator is complex (Särndal et al., 1992, p. 182). There are different approaches to calculate the standard errors asymptotically like linearization and replicate methods, including Jackknife (Tukey, 1958), Bootstrap (Rao & Wu, 1988) or balanced repeated replication estimation (McCarthy, 1966). In this report, the linearization approach, also known as Taylor series estimation or delta method, which is often used in practice, is applied. The main idea is to turn the non-linear estimator into a sum of linear ones using a first-order Taylor approximation of the point estimate of interest (see for example Wolter, 1985, p. 222ff.). The variance of the estimate is then computed using the design-based variance estimator for the total. For a more detailed description and application of the method to different survey estimates see Binder (1983).

Within the class of weighting methods, a lot of different approaches exist: weighting class adjustments (Biemer & Christ, 2008), response propensity modeling (Little, 1986) and calibration methods (Deville & Särndal, 1992) which cover the approaches of poststratification (Valliant, 1993) and raking (Deville et al., 1993). In the following, one of these approaches, the response propensity modeling is described in more detail. The idea of using response propensities for nonresponse adjustments stems from Little (1986) and follows the stochastic concept of the response behavior in which each individual has a non-zero probability of responding which can be denoted by ϕ_i (Kalton & Maligalig, 1991). This probability can be estimated for the sample elements, leading to $\hat{\phi}_i$ which is called *response propensity*. In the application, a logistic regression model for the response indicator is estimated as a function of the available administrative variables.

One important question for modeling response propensities is which of the auxiliary variables at hand should be incorporated in the response propensity model. In this context, it is essential that the included auxiliary variables are available for all initial sample members, respondents and nonrespondents. In general, most researchers in the field (e.g. Groves, 2006; Kalton & Maligalig, 1991; Little, 1986; Little & Vartivarian, 2003, 2005; Peytchev & Olson, 2007; Särndal & Lundström, 2005) declare that nonresponse adjustment will be most effective if variables are used, which are correlated with the probability of responding *and* the key variable of interest. Some authors argue to include as many variables as possible to enhance the predictive power and fit of the model (Heeringa et al., 2010; Krueger & West, 2014) and to make the assumption of MAR more reasonable. However, Kreuter & Olson (2011) show that including variables which are unrelated to the variable of interest may do more harm than good. Using irrelevant information may only inject variability in the estimates without adjusting bias. Moreover, the choice of auxiliary variables has to be made for each Y variable separately since the correlations of the auxiliary and key variables differ. Thus, it is not possible to develop a weighting scheme which is useful for bias reduction in all variables of a survey. In practice, the key problem is that in most surveys only little auxiliary variables (with potentially low correlations to the response indicator and the key variable) are available for both respondents and nonrespondents, limiting the choice of auxiliary variables.

In the following analyses, the auxiliary variables selected in the previous chapter are included in the model. These variables are also used in the introduction of the MAR missing data in the simulated datasets. After estimating the logistic model, the individual response propensities are predicted and used in the adjustments by defining the new weighting factor $w_i = d_i \times \hat{\phi}^{-1}$. The design weights, d_i, are calculated as the inverse of the individual probability to be in the sample which depends on the applied sampling design. Since the samples are drawn by simple random sampling in the simulation study, the design weights are constant and calculated as $N/n = 4276/3000$ (Valliant et al., 2013). For more complex sampling procedures, the determination of the design weights is more difficult and oftentimes they are provided jointly with the survey data.

The response propensities $\hat{\phi}^{-1}$ can be used individually, leading to direct *propensity weighting* and a $\hat{\phi}_i$ in the formula above. If the response model is correct, this procedure enables unbiased or nearly unbiased estimation of population statistics (Heeringa et al., 2010, p. 39). However, problems may arise since logistic models tend to yield estimates which are relatively small, resulting in very large weights and potentially unstable estimates. A possible solution to this is obtained through the creation of classes that are homogeneous with respect to the propensity to respond. One way of forming these classes is the *propensity stratification* approach. Here, the $\hat{\phi}_i$'s are used to create c classes in which all respondents are adjusted with the same factor $\hat{\phi}_c$ (Little, 1986). It follows that the variation between the used $\hat{\phi}$'s will be smaller, leading to more robust estimates and less extreme weights and, hence, a smaller variation in the final estimates. This method is, moreover, less dependent on the accuracy of the response propensity model. The groups are built by sorting the estimated propensities and dividing them into classes with about the same number of units. Following Cochran (1986), the estimated propensities are grouped into five classes. Since the response propensities within each of the five classes show rather little variation, an estimate of the response rate is used as $\hat{\phi}_c$ in each class Valliant et al. (2013, p. 330).

Both approaches, the individual and the stratified versions of the propensity weights, are applied to the data in the estimation of the proportion of homeowners; the results are shown and discussed in chapter 3.2.2.2. Although weighting adjustments may reduce bias due to nonresponse, they do not directly take into account the relationship between response behavior and the survey variables of interest. Therefore, a very different approach to handle missing data is explained in the next chapter.

Bivariate probit model with sample selection

The bivariate probit model with sample selection (Van de Ven & Van Praag, 1981) can be seen as an extension of the Heckman Selection Model (Heckman, 1976, 1979) to model binary outcome variables. The Selection Model was initially developed to determine and correct for specific selection mechanisms. It can, however, also be applied to the situation of survey nonresponse (Cobben, 2009; Nicoletti & Peracchi, 2005), assuming that the sample selection arises due to self-selection of the respondents (by taking part in the survey or not). The model consists of two different equations, the first modeling the probability of response R^* and the second the survey variable of interest Y^*:

$$\begin{aligned} R^* &= \boldsymbol{\beta}_R^T \boldsymbol{x}_R + \epsilon_R && \text{(selection model)} \\ Y^* &= \boldsymbol{\beta}_Y^T \boldsymbol{x}_Y + \epsilon_Y && \text{(outcome model),} \end{aligned} \tag{3.1}$$

where $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_Y$ are vectors of unknown parameters and $\boldsymbol{x}_R$ and $\boldsymbol{x}_Y$ are vectors of auxiliary variables for unit i. The index R is used to flag variables related to the response model, whereas the index Y is added to variables which are modeling the outcome of interest. ϵ_R and ϵ_Y respectively denote the error term for each of the two models. Both of the equations above are latent continuous variable regression equations, representing respectively the propensity to respond and the propensity of Y. R^* and Y^* are not observed, but only the corresponding binary indicators R and Y, with the values $r_i = 1$ if $R^* > 0$ and 0 otherwise and $y_i = 1$ if $Y^* > 0$ and 0 otherwise. Hence, individuals will take part in the survey ($r_i = 1$) if the propensity of responding is high enough ($R^* > 0$ or $\epsilon_R > -\boldsymbol{\beta}_R^T \boldsymbol{x}_R$) and Y will have valid information if the propensity of Y is high enough ($Y^* > 0$ or $\epsilon_Y > \boldsymbol{\beta}_Y^T \boldsymbol{x}_Y$). The sample selection that occurs due to the censoring of the sample elements is introduced by acknowledging that Y is only observed if $r_i = 1$, otherwise Y is missing. To model the correlation between the two equations, the error terms are assumed to be distributed independently of the predictors according to a bivariate standard normal distribution, leading to

$$\begin{pmatrix} \epsilon_R \\ \epsilon_Y \end{pmatrix} \sim NV\left(\mathbf{0}, \begin{pmatrix} 1 & \rho_{RY} \\ \rho_{RY} & 1 \end{pmatrix}\right),$$

where ρ_{RY} denotes the correlation between the two error terms which may arise due to unobserved variables influencing response behavior and the Y variable. If there is no correlation between the two error terms ($\rho_{RY} = 0$), the estimates of the outcome model in equation (3.1) are unbiased. However, since the variable of interest is only partially observed, this assumption may generally not be fulfilled.

The parameters of the model are estimated via full maximum likelihood estimation. In the situation of unit nonresponse, the log-likelihood for the model can be constructed by dividing the sample elements in three mutually exclusive groups, those that are not responding ($r_i = 0$) and, hence, no information about Y is available, those which are responding ($r_i = 1$) and $y_i = 0$ and those respondents with $y_i = 1$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}_R, \boldsymbol{\beta}_Y, \rho_{RY})$ denote the vector of parameters to be estimated jointly. Following Nicoletti & Peracchi (2005), the log-likelihood for the sample of n observations can be written as:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} (1 - r_i)\ ln\ p_{i0}(\boldsymbol{\beta}_R) + r_i\ (1 - y_i)\ ln\ p_{i10}(\boldsymbol{\theta}) + r_i\ y_i\ ln\ p_{i11}(\boldsymbol{\theta})$$

According to Greene (2008), the probabilities underlying the three different realizations of the two binary indicators can be written as

$$\begin{aligned} p_{i0}(\boldsymbol{\beta}_R) &= Pr(r_i = 0 | \boldsymbol{x}_R) = 1 - \boldsymbol{\Phi}(\boldsymbol{\beta}_R^T \boldsymbol{x}_R) \\ p_{i10} &= Pr(r_i = 1, y_i = 0 | \boldsymbol{x}_R, \boldsymbol{x}_Y) = \boldsymbol{\Phi}_2(\boldsymbol{\beta}_R^T \boldsymbol{x}_R, -\boldsymbol{\beta}_Y^T \boldsymbol{x}_Y; -\rho_{RY}) \\ p_{i11} &= Pr(r_i = 1, y_i = 1 | \boldsymbol{x}_R, \boldsymbol{x}_Y) = \boldsymbol{\Phi}_2(\boldsymbol{\beta}_R^T \boldsymbol{x}_R, \boldsymbol{\beta}_Y^T \boldsymbol{x}_Y; \rho_{RY}) \end{aligned}$$

where $\mathbf{\Phi}(\cdot)$ is the standard normal distribution function and $\mathbf{\Phi}_2(\cdot,\cdot;\rho_{RY})$ is the bivariate normal distribution function with zero means, unit variances and correlation coefficient ρ_{RY}.

Similar to the response propensity model, one important decision in this approach is the selection of auxiliary variables in each of the two equations. In general, most research found that the method only works well if both equations contain strong predictors of the respective outcome variable which is problematic since for most cases in practice no detailed information about the two processes is available. In principle, it would be possible to include the same set of covariates in both models. In order to avoid collinearity and to ensure the identification of the model, it is, however, desirable to include at least one variable in the selection model which influences the response process, but not directly the variable of interest. Following these considerations, the selection model in the application is specified equivalently to the response model in the weighting approach, including the selected auxiliary variables plus one additional variable ("change in the unemployment rate") as covariates and the response indicator as the dependent variable. The outcome equation comprises the four auxiliary variables used in the previous approach. The proportion of homeowners is estimated by predicting the missing values based on the results of the outcome model and taking the mean of the predicted probabilities. The resulting estimates are shown and compared with those from the other methods in the next chapter.

3.2.2.2 Estimation results

As already mentioned, T = 100 samples of size n = 3000 are drawn from the initial set of SHARE respondents and, hence, the reported results in this chapter are calculated as the average value over these samples. The estimate for the proportion of homeowners is then derived from $\hat{\hat{Y}} = 1/100 \sum_{t=1}^{100} \hat{\hat{Y}}_t$. The same procedure is applied for the calculation of the bias, variance, and MSE of the estimates. The two methods described in the preceding chapter are applied to the estimation of the proportion and the results for the situation of 60% missing observations are shown in Table 3.2.

	True value	CC	RPW_i	RPW_{str}	BP
$\hat{\bar{Y}}$	0.6368	0.7034	0.5950	0.6459	0.6339
Bias		0.0666	-0.0419	0.0091	-0.0029
Variance		0.0002	0.0005	0.0003	0.0000
MSE		0.0047	0.0028	0.0006	0.0177

Table 3.2: Unadjusted and adjusted estimates of the proportion of homeowners in a scenario with 60% missing observations; "CC"= complete-case, "RPW_i"= individual response propensity weights, "RPW_{str}"= stratified response propensity weights, "BP"= bivariate probit model with sample selection

All of the adjusted estimates do show smaller absolute biases than the complete-case estimate in which only respondents are included in the estimation. Hence, in this scenario the adjustments work well. The estimate of the bivariate probit model with sample selection shows the smallest absolute bias, the average MSE is, however, bigger than those derived from the other methods. The variance is not greatly enhanced due to the application of the adjustment methods. The stratified response propensity method performs best when comparing the MSE.

Since the data situation in this analysis is quite specific, some extensions are applied to the procedure. As a first step, the same analyses are conducted with less missing observations since 60% of missing observations is a rather large amount of missing data and the potential of adjustment methods tends to be decreased when the number of missing values is high (Scheffer, 2002, p. 154; Raymond & Roberts, 1987, p. 20). The amount of missing observations is set to 30% to mirror the target response rate of 70% in the European Social Survey (ESS, Beullens et al., 2016).

	True value	CC	RPW_i	RPW_{str}	BP
$\hat{\bar{Y}}$	0.6368	0.6556	0.6343	0.6406	0.5511
Bias		0.0188	-0.0025	0.0038	-0.0857
Variance		0.0001	0.0001	0.0001	0.0000
MSE		0.0005	0.0002	0.0002	0.0194

Table 3.3: Unadjusted and adjusted estimates of the proportion of the "homeowner" variable in a scenario with 30% missing observations; "CC"= complete-case, "RPW_i"= individual response propensity weights, "RPW_{str}"= stratified response propensity weights, "BP"= bivariate probit model with sample selection

The results in Table 3.3 show that the absolute size of the bias in the estimates is, generally, lower than in the case with more nonrespondents. Moreover, both of the response propensity estimates show smaller absolute values in the bias as the estimate received from the respondent observations only. In this scenario, the individual response propensity approach even outperforms the stratified one due to the smaller bias. However, the bias in the estimate of the bivariate probit model is quite substantial in this case, which might be explained by the lower correlation between the auxiliary variables and the response indicator in this scenario (see Table B.1 in the appendix). This loss of association may lead to lower predictive power of the selection model and, hence, less potential in the estimation of the proportion.

To demonstrate the issue of variable selection more clearly, the analyses with 60% missing observations are additionally run with a modified model construction. For this purpose, the missing data introduction stays the same, but the included auxiliary variables in the adjustments are changed and are, thus, not the same as those used to build the missing data mechanism. For the response propensity model, only one auxiliary variable (% of household size: 2 persons) is used in the logistic model which is only slightly correlated to the response indicator and the homeowner variable in the simulated data. The same modeling applies for the selection equation in the bivariate probit model. The outcome equation in this case only includes age and gender as typical demographic control variables.

	True value	CC	RPW_i	RPW_{str}	BP
$\hat{\bar{Y}}$	0.6368	0.7034	0.7037	0.7032	0.7029
Bias		0.0666	0.0669	0.0664	0.0661
Variance		0.0002	0.0002	0.0002	0.0000
MSE		0.0047	0.0048	0.0047	0.0078

Table 3.4: Unadjusted and adjusted estimates of the proportion of the "homeowner" variable in a scenario with less suitable auxiliary variables; "CC"= complete-case, "RPW_i"= individual response propensity weights, "RPW_{str}"= stratified response propensity weights, "BP"= bivariate probit model with sample selection

The results in Table 3.4 show that the adjusted estimates are worse than in the situation above, in which comparably good auxiliary variables are used in the adjustments. All estimates are rather close to the unadjusted one regarding the absolute size of bias. These results clearly demonstrate the sensitivity of the potential of the methods which highly depends on the included auxiliary variables.

Summarizing the results regarding the estimate of a proportion, it can be stated that nonresponse bias does exist in the MAR missing data specification applied and, as such, simple ad-hoc methods may not be appropriate for the estimation. We thus tested the performance of two response propensity approaches and both perform well. They substantially reduce the arising bias in comparison to the complete-case estimates. The bivariate probit model with sample selection is in this case also able to correct for bias, the potential of the method is, however, limited in practice since it requires strong predictors in both estimation equations. In general, it must be emphasized that all of the observed methods probably work well in this data situation since detailed knowledge of the influencing factors of the response process is available and the missing data are missing at random. In practice, this information will often be unavailable, limiting the potential of the adjustment methods (as was shown in the analysis with less suitable auxiliary variables). Moreover, the observed correlations between the indicators of interest are quite high on average and it is uncertain whether auxiliary variables with such high associations can be found in practice. Furthermore, in many missing data situations, researchers might have to deal with data where observations are not missing at random, which in most cases lead to biases in the estimates that are not correctable by our adjustment methods. It has also been shown that the occurring bias is stronger, the higher the nonresponse rate is. This points to the relevance of high response rates in social surveys even if a higher response rate does not necessarily result in smaller nonresponse errors (Groves & Peytcheva, 2008). In summary, adjustments based on the auxiliary regional information are indeed able to reduce bias in univariate estimates. However, the performance depends on specific features in the data which are hard to determine in practice.

After analyzing possible nonresponse reduction in one specific data situation, the question remains whether there are, in general, useful auxiliary variables for nonresponse analysis also at cross-national level. For this purpose, the next chapter gives an overview of previous findings in this context.

4 Auxiliary data sources

Although based on only one country, the previous section nevertheless made clear that auxiliary data can help to reduce nonresponse bias - at least if certain conditions are fulfilled. The purpose of this section is hence to give an overview of potential data sources for auxiliary variables and prior work in this research field in order to be able to better evaluate if and how the proposed approach can be extended to other (cross-national) surveys within the SERISS project. In this respect, the following will add further information on existing knowledge both from previous research (see below) and from cross-national studies, such as the GGP that already has some experience with administrative data in certain countries (Caporali et al., 2016). Taken together, this should help to be able to better evaluate the potential of auxiliary data for nonresponse bias adjustments.

Generally, multiple sources of auxiliary variables are available. This includes paradata, like contact protocols and disposition codes, collected for all survey units (Couper, 1998; Kreuter & Casas-Cordero, 2010), interviewer observations (e.g. Blom et al., 2010; Campanelli & O'Muircheartaigh, 1999; Durrant et al., 1997; Groves & Couper, 1998; Groves et al., 2007; Lipps & Benson, 2005; Peytchev & Olson, 2007) or other indicators from external databases (Smith & Kim, 2013). However, it is important to consider the quality of the information, if the data are accessible for research purposes and whether the available variables are useful in nonresponse bias adjustment with regard to the criteria mentioned earlier. The role of interviewers in collecting auxiliary information has recently received much attention in the survey methodology literature (see for example Groves & Couper, 1998) because interviewer observations (for example regarding the area or the building respondents are living in) can be easily obtained during survey data collection. However, it has been shown that interviewer observations are subject to the interpretation of the interviewer and can, thus, introduce measurement error and additional variance into the data (Kaminska & Lynn, 2011; Sinibaldi et al., 2011; West, 2013).

The main part of this report (also the scoping study described earlier) focuses on external data sources. Information on dwellings and neighborhoods of potential respondents have been combined with survey data to analyze response behavior in more detail (Groves & Couper, 1998; House & Wolf, 1978; Schräpler et al., 2010; Steeh, 1981). In fact, geographical areas have been identified as a source of potential variation in response rates (see for example Hox & De Leeuw, 2002) and may, thus, be useful for nonresponse bias adjustment. The advantages of using this kind of data lie in the reduced costs and respondent burden since no additional questions are necessary during fieldwork. Moreover, since most of the indicators are released on an aggregate level only, the information is relatively easy accessible and oftentimes free of charge, and no respondent consent to link survey responses to administrative records is needed. This is beneficial since it has been shown that asking for consent often leads to a strongly reduced sample size and refusal may also introduce bias (Kreuter et al., 2010a). In addition, small-area variables have been shown to be of good quality and contain little missing data, making them a valuable source for additional variables (Butt & Lahtinen, 2016).

From the literature about possible sources of auxiliary variables, two prior projects, which examine various sources of external auxiliary variables, are of particular importance: First, Deliverable 2.5 of the SERISS project (Bristle et al., 2016) gives an overview of pre-existing administrative information from different countries that can potentially be used as auxiliary variables. The aim of this project was to provide an inventory of auxiliary data that are available in registers used as sampling frames in the four surveys participating in the project (ESS, EVS, GGP, and SHARE). The information from country registers and, in addition, data sources on the European level (from Eurostat or Census Hub) was evaluated across three dimensions: the availability, the possibility of accessing the data, and the geographical identifiers for linking both data sources. The advantage of information derived from country registers is that the data are available for all sample units whereas the potential of these data is limited through the content and possible missing data. In comparison, information from the European data sources offers a wide range of indicators from various areas, but the data are most often only available on an aggregated level. The results of this project are summarized as follows: most observed registers do contain auxiliary information, but the availability is limited to general socio-demographic characteristics like age and gender. Household characteristics are included less often and information about the neighborhood is very rarely given. Furthermore, the geographical identifiers included in the data often differ regarding their regional level and the codes can rarely be used for matching in practice. The potential for nonresponse bias reduction of these auxiliary variables may be bounded by these limitations. Moreover, data availability and access conditions vary significantly across countries, surveys, and organizations within countries, which complicates the possibility of cross-national usability of single variables.

The second major project with regard to the collection and evaluation of auxiliary variables for nonresponse analysis is the ADDResponse project (Butt & Lahtinen, 2016), which was launched to evaluate the availability and suitability of geocoded data for appending address-based survey samples in the UK. Therefore, auxiliary data from three external sources – small-area administrative data, geocoded information on the location of the sample units and commercial marketing data – were appended to the UK sample of the ESS in Round 6 (2012/ 2013). The various data sources are evaluated regarding their quality, accessibility, and the associated costs, allowing several conclusions to be drawn. First, there is a growing amount of auxiliary data available to survey researchers, which can be linked to survey data at address or municipality level without the need to obtain specific consent. The ADDResponse project was able to append around 400 auxiliary variables from 20 different data sources to the survey sample. Second, much of the observed data are freely available without restrictions. Other data may only be available for a limited time period, depending on the purpose of data usage. Third, auxiliary data should be accessed and appended to the sample at the time the sample is drawn, prior to fieldwork. Fourth, data access for the purpose of appending additional information to the survey data is relatively unproblematic, but sharing the combined data is less straightforward since access to the built dataset of ADDResponse is generally restricted to the research team. Fifth, many data are country-specific and, as such, the availability, definition, and measurement as well as the coverage vary across countries. With regard to the potential of the linked variables for adjusting nonresponse bias, the authors conclude that despite the large amount of collected information it is difficult to identify variables which are significantly correlated to the response propensity and substantive survey variables.

Against this background, it has to be concluded that there are considerable difficulties of identifying similar auxiliary variables in a cross-national perspective. In addition, the different SERISS surveys focus on very different types of questions (e.g. questions on the economic or health situation of respondents in SHARE vs. attitudes in ESS). This may mean that different auxiliary variables are required for ESS, EVS, GGP, and SHARE. In this respect, it is hardly surprising that there is an ongoing discussion in several SERISS surveys about the advantages and disadvantages of using different auxiliary variables to construct weights in different countries. Based on the results of the simulation study, there is evidence that optimizing country-specific models might be advantageous - at least if there are variables that fulfill the criteria mentioned above.

5 Summary & Conclusion

The main objective of this report was to provide an overview of available auxiliary variables and their potential for nonresponse bias adjustment. Generally, a wide range of auxiliary variables is available from various sources including paradata, interviewer observations from the survey process or contextual information from external data sources like registers or Censuses. The accessibility and usefulness of the different indicators, however, vary across measures, countries, and institutions. This variability makes it difficult to build a joint database or give general recommendations.

It must be emphasized that nonresponse bias analysis and the evaluation of the potential of different adjustment methods based on survey data alone is difficult since the true population values are unknown and the bias cannot be quantified. To overcome this limitation, we simulated a dataset based on real survey data where the reference value is known. This approach enables the quantification of nonresponse bias and the comparison of the potential of various methods for reducing this bias. Regional information from two different external sources in Germany has been selected as one specific set of administrative variables. Two adjustment methods based on these data were then applied to the estimation of the proportion of one specific survey variable. We used one quite common weighting method, response propensity weighting, and a maximum-likelihood based method, the bivariate probit model with sample selection. The results show that complete-case estimates which ignore missing observations can exhibit high biases from the true population value. Adjustments based on auxiliary variables, which are highly correlated with the probability to respond as well as the survey item of interest, are able to reduce the absolute bias in the estimate if the missing information is missing at random (MAR). The potential of the adjustments, however, heavily relied on the quality of the auxiliary variables. Moreover, it has been shown that higher survey nonresponse rates tend to result in more strongly biased estimates.

Concluding from the results of this report and various other research projects regarding the reduction of nonresponse bias, the most important task for analysts is the identification of suitable auxiliary variables which are associated with both the response probability and the survey variable of interest. In practice, this can be difficult since many auxiliary variables are quite specific and may be unrelated or only weakly related to survey items. The two prior research projects presented in chapter 4 (ADDResponse, Butt & Lahtinen (2016) and SERISS D2.5, Bristle et al. (2016)) showed that a lot of auxiliary variables most likely do not exhibit strong associations with the response indicator and the survey variables of interest. This may either be due to the limited information available as register information is often limited to demographic variables or due to the aggregated level of the auxiliary indicators which are available from other regional data sources like Censuses.

Another possibility for obtaining auxiliary variables of good quality which could be applied in future surveys is the basic question method proposed by Bethlehem & Kersten (1985). The idea behind this approach is that individuals who refuse to respond to the whole survey may be willing to answer one or two basic questions. Such questions could, for example, ask about social activities, civil engagement and attitudes towards surveys and move beyond demographic characteristics. In the best case, these variables should be associated with the central survey items and the response propensity and could then be used for adjustments. While being applied in single countries in the ESS (see Matsuo et al., 2010, for further information and empirical findings), this method has still to be tested in a cross-national environment. However, with ex-ante harmonization of such questionnaires for nonrespondents, it should, in principle, be possible to obtain comparable and valuable information. On the other hand, usually not all nonrespondents take part in follow-up surveys, which reduces their utility in practice.

Future research in the field of nonresponse bias adjustment might take several directions. While the search for well-fitting auxiliary variables in registers and on regional level can be tedious, they have the potential to reduce the observed bias under the missing at random condition and acceptable size of nonresponse rate. Making the effort to ask nonrespondents about characteristics closely related to nonresponse and to the content of the survey might be a promising extension.

A Auxiliary variables overview

Variable name	Description	Source
individual		
age	age of the sample unit	S
female	gender of the sample unit	S
lau2		
gk_pol	political municipality size (in five categories)	C
pop_male	percentage of male population	C
fam_sim	percentage of marital status: single	C
fam_mar	percentage of marital status: married	C
fam_wid	percentage of marital status: widowed	C
fam_div	percentage of marital status: divorced	C
agecat_9	percentage of population aged < 9 yrs	C
agecat_1019	percentage of population aged 10 - 19 yrs	C
agecat_2029	percentage of population aged 20 - 29 yrs	C
agecat_3039	percentage of population aged 30 - 39 yrs	C
agecat_4049	percentage of population aged 40 - 49 yrs	C
agecat_5059	percentage of population aged 50 - 59 yrs	C
agecat_6069	percentage of population aged 60 - 69 yrs	C
agecat_7079	percentage of population aged 70 - 79 yrs	C
agecat_80	percentage of population aged > 80 yrs	C
nat_ger	percentage of German nationality	C
bcnt_ger	percentage of birth country Germany	C
rel_cat	percentage of Catholics	C
rel_pro	percentage of Protestants	C
rel_other	percentage of other religious affiliation	C
use_ibo	percentage of houses inhabited by owner	C
hhtype_sin	percentage of household type: single household	C
hhtype_mar	percentage of household type: married couple	C
hhtype_coh	percentage of household type: cohabiting union	C
hhsize_2	percentage of household size: 2 persons	C
hhsize_3	percentage of household size: 3 persons	C

hhsize_4	percentage of household size: 4 persons	C
hhsize_5	percentage of household size: 5 persons	C
hh_old	percentage of households with senior citizens only	C
hh_med	percentage of households with senior and younger citizens	C
hh_you	percentage of households with no senior citizens	C
unemp	unemployment rate	I
unemp_change	change in unemployment rate (last 5 years)	I
emp	employment rate	I
mig_bal	migration balance	I
pop_den	population density (inhabitants per km^2)	I
nuts3		
unemp_old	unemployment rate elderly citizens (aged 55+)	I
emp_old	percentage of employed elderly citizens (aged 55+)	I
min_emp_old	percentage of elderly citizens (aged 65+) in minor employment	I
med_age	medium age in population	I
hhinc	household income	I
medinc	median income	I
med_pens	medium pension fee	I
numb_psych	number of psychotherapists per 100000 inhabitants	I
numb_hospbed	number of hospital beds per 10000 inhabitants	I
numNOC	number of persons in need of care per 10000 inhabitants	I
finsupp	percentage of persons in need of care receiving financial support	I
per_poorold	percentage of poverty among elderly citizens (aged 65+)	I
gdp	gross domestic product per inhabitant	I

Table A.1: Overview of available auxiliary variables with name, definition, and source (C=Census, I=INKAR, S=SHARE), “lau2”: information on municipality level, “nuts3”: information on county level

B Tables & Figures

B.1 Correlation analysis in the original SHARE data

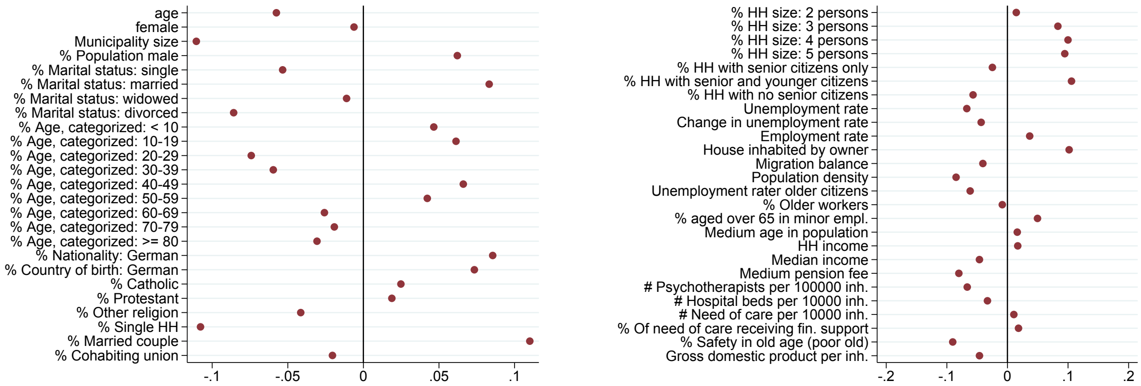

Figure B.1: Correlations of auxiliary variables (X) with the response indicator (R) in the original SHARE data

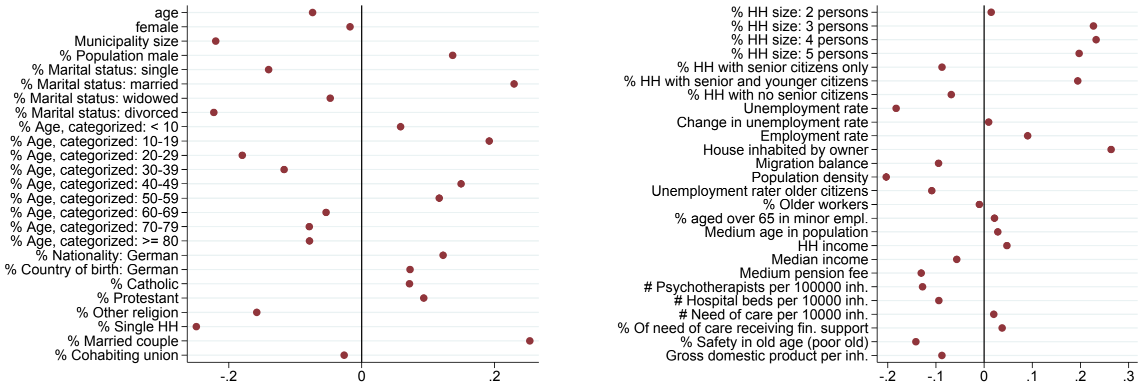

Figure B.2: Correlations of auxiliary variables (X) with the survey variable "homeowner" in the original SHARE data

B.2 Correlation analysis in the simulated data with 30% missing observations

Auxiliary variable	Correlation with R	Correlation with Y
Municipality size	-0.2183	-0.2229
% houses inhabited by owner	0.2118	0.2572
% household type: married couple	0.2307	0.2515
% households with senior and younger citizens	0.2792	0.1866

Table B.1: Average correlations between the auxiliary variables and the response indicator and the "homeowner" variable in the 100 simulated datasets with only 30% missing observations

Bibliography

Bethlehem, J. G., & Kersten, H. M. P. (1985). On the treatment of nonresponse in sample surveys. *Journal of Official Statistics*, *1*, 287–300.

Beullens, K., Loosveldt, G., Denies, K., & Vandenplas, C. (2016). Quality matrix for the european social survey, round 7.

Biemer, P. P., & Christ, S. L. (2008). Weighting survey data. In J. E. Hox, E. D. Leeuw, & D. Dillman (Eds.) *International handbook in survey methodology*, (pp. 317–341). London: Lawrence Erlbaum Publishers.

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, *51*(3), 279–292.

Blom, A. G., De Leeuw, E., & Hox, J. J. (2010). Interviewer effects on nonresponse in the European Social Survey. *MEA Discussion Papers*, *202-2010*.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., & Zuber, S. (2013). Data resource profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*.

Brick, A. M., & Williams, D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. *The Annals of the American Academy of Political and Social Science*, *645*(1), 36–59.

Bristle, J., Butt, S., Emery, T., Luijx, R., Maineri, A. M., Pflüger, S.-M., Scherpenzeel, A., & Zins, S. (2016). Report on auxiliary data in available country registers. deliverable 2.5 of the seriss project funded under the european union's horizon 2020 research and innovation programme ga no: 654221.

Butt, S., & Lahtinen, K. (2016). Addresponse: Auxiliary data driven nonresponse bias analysis - technical report on appending geocoded auxiliary data to round 6 of european social survey (uk).

Campanelli, P., & O'Muircheartaigh, C. (1999). Interviewers, interviewer continuity, and panel survey nonresponse. *Quality & Quantity*, *33*, 59–76.

Caporali, A., Neyer, G., Grigorieva, O., Klüsener, S., Krapf, S., & Kostova, D. (2016). The contextual database of the Generations and Gender Programme: Concept, content, and research examples. *Demographic Research*, *35*(9), 229–252.

Cobben, F. (2009). *Nonresponse in sample surveys – methods for analysis and adjustment*. PhD thesis, University of Amsterdam, The Hague.

Cochran, W. (1986). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, *24*, 295–313.

Collins, L. M., Schafer, J. L., & C.-M., K. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351.

Couper, M. (1998). Measuring survey quality in a CASIC environment. In *Proceedings of the Section on Survey Research Methods of the American Statistical Association*.

De Heer, W., & De Leeuw, E. (2002). Trends in household survey nonresponse. *Survey Nonresponse*, *41*.

Deville, J.-C., & Särndal (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, *87*, 376–382.

Deville, J.-C., Särndal, C.-E., & Sautory, O. (1993). Generalizied raking procedures in survey sampling. *Journal of the American Statistical Association*, *88*, 1013–1020.

Durrant, G. B., Groves, R. M., Staetsky, L., & Steele, F. (1997). Effects of interviewer attitudes and behaviors on refusal in household surveys. *American Journal of Epidemiology*, *146*(5), 439–449.

Glynn, R. J., Laird, N. M., & Rubin, D. B. (1993). Multiple imputation in mixture models for nonignorable nonresponse with followups. *Journal of the American Statistical Association*, *88*, 984–993.

Graham, J. W., & Donaldson, S. I. (1993). Evaluating interventions with differential attrition: the importance of nonresponse mechanisms and use of follow-up data. *Journal of Applied Psychology*, *78*, 119–128.

Greene, W. H. (2008). *Econometric analysis*. Pearson Education.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646–675.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. John Wiley & Sons.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*. Hoboken, NJ: John Wiley & Sons.

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

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. *Public Opinion Quarterly*, *72*(2), 167–189.

Groves, R. M., Wagner, J., & Peytcheva, E. (2007). Use of interviewer judgements about attributes of selected respondents in post-survey adjustment for unit nonresponse: an illustration with the National Survey of Family Growth. In *Proceedings of the Survey Research Methods Section*, (pp. 3428–3431).

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economics and Social Measurement*, *5*(4), 475–492.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrics*, *47*(1), 153–161.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). *Applied survey data analysis*. Chapman & Hall.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*(260), 663–685.

House, J., & Wolf, S. (1978). Effects of urban residence and interpersonal trust and helping behavior. *Journal of Personality and Social Psychology*, *10*, 222–226.

Hox, J. J., & De Leeuw, F. (2002). The influene of interviewer's attitude and behaviour on household survey nonresponse: an international comparison. In R. M. Groves, A. Dillman, J. L. Eltlinge, & J. A. Little (Eds.) *Survey nonresponse*. New York: Wiley.

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

Kaminska, O., & Lynn, P. (2011). Interviewer observations: are the observations what we want or what interviewers think we want? *Paper presented at the 22nd International Workshop on Household Survey Nonresponse*.

Kish, L. (1965). *Survey sampling*. New York: John Wiley & Sons.

Kish, L. (1992). Weighting for unequal pi. *Journal of Official Statistics*, *8*, 183–200.

Kreuter, F., & Casas-Cordero, C. (2010). Paradata. *Working Paper Series des Rates für Sozial- und Wirtschaftsdaten*, *136*.

Kreuter, F., Müller, G., & Trappmann, M. (2010a). Nonresponse and measurement error in employment research – making use of administrative data. *Public Opinion Quarterly*, *74*(5), 880–906.

Kreuter, F., Olson, J., Wagner, T., Yan, T., Ezzati-Rice, T. M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R. M., & Raghunathan, T. E. (2010b). Using proxy measures and other correlates of survey outcomes to adjust for nonresponse: examples from multiple surveys. *Journal of the Royal Statistical Society Series A*, *173*(2), 389–407.

Kreuter, F., & Olson, K. (2011). Multiple auxiliary variables in nonresponse adjustment. *Sociological Methods & Research*, *40*(2), 311–332.

Krueger, B., & West, B. (2014). Assessing the potential of paradata and other auxiliary data for nonresponse adjustments. *Public Opinion Quarterly*, *78*(4), 795–831.

Lepkowski, J. M., & Couper, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. In R. M. Groves, A. Dillman, J. L. Eltlinge, & J. A. Little (Eds.) *Survey nonresponse*. New York: Wiley.

Lessler, J., & Kalsbeek, W. D. (1992). *Nonsampling error in surveys*. New York: John Wiley and Sons.

Lipps, O., & Benson, G. (2005). Cross-national contact strategies. In *Proceedings of the Survey Research Section of the American Statistical Association*. Alexandria.

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

Little, R., & Vartivarian, S. (2003). On weighting the rates in non-response weights. *Statistics in Medicine*, *22*(9), 1589–1599.

Little, R. J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Mathematica Policy Research*.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. John Wiley & Sons, 2nd ed.

Matsuo, H., Billiet, J., Loosveldt, G., Berglund, F., & Kleven, Ø. (2010). Measurement and adjustment of non-response bias based on non-response surveys: the case of belgium and norway in the european social survey round 3. *Survey Research Methods*, *4*(3), 165–178.

McCarthy, P. J. (1966). Replication: an approach to the analysis of data from complex surveys. *Vital and Health Statistics*, *2*.

Nicoletti, C., & Peracchi, F. (2005). Survey response and survey characteristics: microlevel evidence from the European Community Household Panel. *Journal of the Royal Statistical Society: Series A*, *168*(4), 763–781.

Peytchev, A., & Olson, K. (2007). Using interviewer observations to improve nonresponse adjustments: NES 2004. *Sociology Department, Faculty Publications*, *145*.

Rao, J. N. K., & Wu, C. F. J. (1988). Resampling inference with complex survey data. *American Statistical Association*, *83*, 231–241.

Raymond, M. R., & Roberts, D. (1987). A comparison of methods for treating incomplete data in selection research. *Educational and Psychological Measurement*, *47*, 13–26.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Särndal, C.-E., & Lundström, S. (2005). *Estimation in surveys with nonresponse*. John Wiley & Sons.

Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. Springer.

Scheffer, J. (2002). Dealing with missing data. *Research Letters in the Information and Mathematical Sciences*, *3*, 153–160.

Schouten, B., Calinescu, M., & Luiten, A. (2013). Optimizing quality of response through adaptive survey design. *Survey Methodology*, *39*, 29–58.

Schräpler, J.-P., Schupp, J., & Wagner, G. (2010). Individual and neighborhood determinants of survey nonresponse: an analysis based on a new subsample of the German Socio-Economic Panel (SOEP), microgeographic characteristics and survey-based interviewer characteristics. *SOEPpapers on Multidisciplinary Panel Data Research*, *288*.

Singer, E. (2002). Use of incentives to reduce nonresponse in household surveys. In R. M. Groves, A. Dillman, J. L. Eltlinge, & J. A. Little (Eds.) *Survey nonresponse*. New York: Wiley.

Singer, E. (2006). Introduction: nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 637–645.

Sinibaldi, J., Kreuter, F., & Durrant, G. (2011). Evaluating the measurement error in interviewer observed paradata. *Paper presented at the 22nd International Workshop on Household Survey Nonresponse.*

Smith, T. W., & Kim, J. (2013). An assessment of the Multi-level Integrated Database approach. *Annals of the American Academy of Political and Social Science*, *645*, 185–221.

Steeh, C. (1981). Trends in nonresponse rates. *Public Opinion Quarterly*, *45*(1), 40–57.

Tukey, J. W. (1958). Bias and confidence in not-quite large samples. *Annals of Mathematical Statistics*, *29*, 614.

Valliant, R. (1993). Poststratification and conditional variance estimation. *Journal of the American Statistical Association*, *88*, 89–96.

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples.* New York: Springer.

Van de Ven, W., & Van Praag, B. (1981). The demand of deductibles in private health insurance: a probit model with sample selection. *Journal of Econometrics*, *17*(2), 229–252.

West, B. (2013). An examination of the quality and utility of interviewer observations in the National Survey of Family Growth. *Journal of the Royal Statistical Society: Series A*, *176*(1), 211–225.

West, B. T., & Little, J. A. (2013). Nonresponse adjustment of survey estimates based on auxiliary variables subject to error. *Journal of the Royal Statistical Society, Series C*, *62*, 213–231.

Wolter, K. M. (1985). *Introduction to variance estimation.* New York: Springer Verlag.