How to make internet surveys representative: A case study of a two-step weighting procedure*

Axel Börsch-Supan
University of Mannheim

Detlev Elsner
McKinsey & Company

Heino Faßbender
McKinsey & Company

Rainer Kiefer
McKinsey & Company

Daniel McFadden
University of California, Berkeley

Joachim Winter**
University of California, Berkeley, and University of Mannheim

This version: March 30, 2004

* The authors appear in alphabetical order. We would like to thank Alexandru Degeratu, Ralph Fries, Hans Gersbach, Titus Kehrmann, Michael Kliger, Gerhard Schulze, and Claudia Stachel for useful comments and support. The usual disclaimer applies.

** Corresponding author.
Address: Department of Economics, University of Mannheim, D-68131 Mannheim, Germany.
E-mail: winter@econ.berkeley.edu

How to make internet surveys representative: A case study of a two-step weighting procedure

Abstract: Internet surveys are becoming increasingly popular. Concerns about the representativeness of online samples, however, frequently cast doubts on the validity of conclusions derived from internet survey data. These doubts rest on the fact that not all persons have internet access and on the premise that people are more likely to participate in a survey if the subject matter interests them. This paper presents a two-step procedure for weighting data from online surveys that is based on an explicit behavioral model of internet access and survey participation decisions. We illustrate the application of this scheme in a case study of *Perspektive Deutschland*, a very large online survey that was conducted in Germany in 2001 with almost 170,000 online respondents. We discuss practical aspects of weighting procedures such as problems caused by large weights as well as theoretical aspects such as the statistical implications of weighting.

Keywords: internet surveys; sample selection; weighting

1. Introduction

Internet surveys are becoming increasingly popular. Various branches of applied statistics begin to rely on data from online samples, such as opinion polls (e.g., Taylor et al., 2001), marketing and business research (e.g., Clayton and Werking, 1998; McCullough, 1998), and research in the social sciences more generally (e.g., Couper, 2000; Schonlau et al., 2002). Concerns about the representativeness of online samples, however, frequently cast doubts on the validity of conclusions derived from internet survey data. These doubts rest on the fact that not all persons have internet access and on the premise that people are more likely to participate in a survey if the subject matter interests them. This paper addresses the problem of representativeness of online surveys. We present a two-step procedure for weighting data from online surveys that is based on an explicit behavioral model of internet access and survey participation decisions. This approach has been developed for *Perspektive Deutschland* (which may be translated as "Germany: Looking Ahead"), a large-scale online survey conducted in Germany in 2001, 2002, and 2003. The present paper uses data from the 2001 wave with some 170,000 online participants.

Various ways to derive weights for online surveys have been discussed in the literature. An extensive overview of statistical approaches for weighting and poststratification can be found in Kalton and Flores-Cervantes (2003). A method that is commonly used is to correct for sample selection in online surveys is propensity score weighting, see Schonlau et al. (2004) for a discussion. The two-step weighting approach we present in this paper is akin to propensity score weighting. The main innovation of our approach is that it is based on an explicit behavioral model of how participants self-select into an online sample. This model reflects the observation that self-selection into an online sample is the joint effect of two factors: internet access and willingness to participate (see, for example, Bosnjak et al., 2001). Our approach of spelling out an explicit model of participation is routed in the econometric literature on sample selection and choice-based sampling (Heckman, 1976; McFadden et al., 1977; Manski and Lerman, 1977; Coslett, 1981; McFadden, 2001).[1]

Our model starts from the observation that in the population, access to the internet differs across socio-demographic groups. Table 1 shows internet access rates by age, gender, and education for both Germany and the U.S. While the U.S. has higher rates of internet access, the overall picture is identical: People who have access to the internet are younger and more educated than the population average, and men have better access to the internet than women. The numbers shown in table 1 show clearly that online surveys are, generally, not representative of the population because of differential internet access. This problem is addressed by the first component of our behavioral model, the internet access model.

Even after removing selection that results from differential access to the internet, a weighted online sample may not be representative of the population. Even conditional on having access to the internet, not all people are reached by various forms of solicitation to participate with the same probability. Conditional on being exposed to some form of solicitation, the decision to participate might still depend on personal characteristics which are – a most important point – correlated with the subject matter which we would like to assess in the survey. Surveys on automobile usage will attract more attention and participation by automobile enthusiasts, and surveys about political responsibility more by politically responsible persons. Answers to the internet survey questions will therefore be biased since we have too many auto enthusiasts (or politically responsible persons) in our sample. This kind of bias is particularly dangerous because it tends to confirm our beliefs. In the language of econometrics, the bias is particularly dangerous because the participation decision is endogenous with respect to the features which we try to measure.

The application of our two-step weighting approach is illustrated in a case study of *Perspektive Deutschland*, a large online survey that has been conducted in Germany in three waves since 2001. This survey is ideally suited to show the strength and weaknesses of

---

[1] The present paper is intended as a case study of weighting and therefore does not exploit the full potential of the behavioral model of internet survey participation. Extensions to a fully developed econometric framework of choice-based sampling which combines and jointly estimates models of participation and response (such as McFadden et al., 1985, and Hellerstein and Imbens, 1999) would be straightforward, but is beyond the scope of this paper.

internet surveys, and, most importantly, how to mitigate the weaknesses. The survey was designed to ascertain opinions about social and political topics, and to measure the willingness to do something about solving social and political problems. The first wave of the survey drew a large response from almost 170,000 respondents in just two and a half months in late 2001.[2] The large number of participants gave the survey enormous political weight, even though the sheer number of participants in a selected sample does of course not ensure that a survey's results are informative. Both selectivity problems – a sample of internet users, and a highly political survey content – had to be tackled in order to create confidence in the survey results. The present paper describes a practical way to address these problems, and it can be viewed as a case study of the strengths and weaknesses of internet surveys more generally.

The structure of the paper is as follows. Section 2 describes *Perspektive Deutschland* in order to set the stage. In Section 3, we present our two-step model of online survey participation and the corresponding weighting procedure. Section 4 describes how this approach was implemented in the analysis of *Perspektive Deutschland*. We illustrate the results of the weighting procedure by comparing weighted online and offline responses. Section 5 addresses practical issues related to large weights. In section 6, we summarize our findings and discuss implications for the application of internet surveys in marketing research and other fields.

2. Our case study: *Perspektive Deutschland*

The objective of *Perspektive Deutschland* was to provide the German public with a forum to voice their opinions about social and political topics. The survey was a joint initiative by McKinsey, a management consulting firm, stern.de, the internet version of a large-circulation weekly magazine, and T-Online, the largest internet survey provider in Germany. The survey was carried out from October 18 to December 31, 2001. With almost 170,000 respondents, *Perspektive Deutschland* was one of the largest online surveys administered at that time; the subsequent waves attracted even larger numbers of participants in 2002 and 2003. The same

[2] The subsequent waves had even larger numbers of participants – about 350,000 in 2002 and 450,000 in 2003.

two-step weighting method was used for the 2002 and 2003 waves. In the sequel, we focus on the 2001 data.

The outstanding number of participants of *Perspektive Deutschland* roots in a broad media campaign. Numerous advertisements in print media and the internet communicated the survey as a way for Germans to express their opinion about economic reforms that would address the country's pressing growth and unemployment problems. The media campaign stressed that the results would be broadly communicated, would be brought to the attention of key decision makers, and would be used to spark off a public debate.[3] Therefore, we believe, many participants saw the survey as an unique opportunity to express their opinion about the most pressing public policy issues in Germany.

The online survey was split into a core block that was presented to all participants and three theme blocks that were randomly assigned to participants. The theme blocks covered three areas: (1) work and leisure, (2) education, and (3) savings, retirement planning, and insurance. The average completion time for the core block was 21 minutes; average completion times for the theme blocks were as follows: work and leisure, 7 minutes; education, 9 minutes; and savings and insurance, 7 minutes.

The random-block design was chosen for two reasons. First, this design reduces average completion time for the online survey. The design had set a target average completion time of 30 minutes to avoid break-offs and drop-outs (which are known to be a function of survey time). Second, the random-block design reduces the risk of strategic self-selection into blocks that are of special interest to participants since each block is presented only with a probability of 1/3. The disadvantage of this design is that the number of observations is significantly reduced when the weighting scheme uses variables from the theme blocks.

The core block included general questions: how is the quality of your life in general, how satisfied are you with the city or community of residence, what are Germany's most pressing

[3] In total 144 different access channels were used. Most of our participants were recruited via e-mail newsletters. The response rate was highest if our survey was mentioned in the subject of the newsletter or was the first topic of the newsletter. The channels also differ in their effectiveness as measured by the completion ratio (the ratio of participants who completed the survey to participants who started the survey). With respect to completion, the worst channels were banners on web pages and the best were press reports.

economic problems, do you want more or less state intervention, which matters should be managed by the state and which privately, etc., and the usual array of socio-demographic characteristics such as age, marital status, education, and income class. It also contained questions about civil engagement such as the willingness to bear reforms even if they incur short-run costs to the respondent.

The theme block on work and leisure focussed on job characteristics (would you like more flexible work hours, would you like to work more/less, would you consider becoming self-employed) and labor market policy (should unemployed become more/less state support, should unemployment compensation be limited in time). The theme block on education addressed willingness to pay for education, willingness to undergo life long learning programs, and the most important content areas of education. Finally, the theme block on savings, retirement planning, and insurance tried to ascertain wealth, savings motives, expected retirement age, and preferences about own provision for retirement income.

Some of the results were expected – for instance, the large gap in the quality of life between the former East and West Germany – but the survey also uncovered some important new insights. Most striking was the large proportion of respondents willing to reform Germany's social and political system. For instance, the results from *Perspektive Deutschland* suggest that Germans may be more ready than their politicians for economic reforms that could solve some of the country's economics problems such as high unemployment and unsustainable pensions. Germans also seem to be more willing to contemplate lower levels of government support than politicians generally acknowledge. For instance, in a country accustomed to free university education, only about a third of the respondents said that they would not under any circumstances pay part of the cost of improving the system. A larger group – about 40 percent – would pay tuition if low-interest loans were available, while an additional 28 percent would pay a part of the cost of higher education if their money could improve it in a noticeable way. These and many other results of *Perspektive Deutschland* received considerable media attention, not the least because of the large number online participants.

While the large number of participants in this online survey increased its weight in the public debate, it did obviously not solve the problem that such an online survey is not representative of the population. The implementation of a weighting procedure ensuring that final results are representative was therefore a central part of the *Perspektive Deutschland* project.

The construction of weights of course needs a gauge. The gold standard gauge are population statistics from the German *Mikrozensus* (a micro census conducted by the Federal Statistical Office that covers about one percent of the population). We used *Mikozensus* data wherever possible, in particular for basic socio-demographic charactersitics. However, many variables that are related to participation in a specific survey are not contained in multi-purpose population surveys such as the *Mikrozensus* – for instance, variables such as political interest. Hence, we combined our (large) online survey with a (small) traditional CAPI survey that used a random sample of the adult population. In the sequel, we refer to this second survey as the "offline survey". This parallel study was administered with a random-route sampling protocol and a total of 2,715 participants.[4] The offline survey used the same questions as the online survey.[5] The data from the offline survey, together with *Mikrozensus* data, were used to construct weights in order to make the results from the online survey representative of the population.

We should note that the data required to construct weights such as political could potentially be obtained from other surveys. This approach would, however, create some problems. First, data from several sources would have to be combined; in our approach, we rely only on two sources. Second, some variables that predict participation in a survey motivated by topical political debates may not be stable over time so that the data used for constructing weight should ideally be collected during the field period of the main survey. Finally, a parallel offline survey has the additional advantage that the distributions of response variables in the weighted online and the offline samples can be compared to evaluate the success of the weighting procedure. We discuss instances in which it may not necessary to conduct a parallel offline survey in the concluding section. Whether a parallel offline survey is conducted or not raises, of course, cost issues – the additional cost may even outweigh the cost advantages of to online survey.

An advantage of the large number of online participants is that the weighted online data can be analyzed at the regional level, assuming that the selection process into the online survey is

[4] In the remainder of this paper, we refer to this parallel study as the "offline survey"

[5] Although the questions were identical, there were some differences in administration between the online and offline surveys, discussed in section 3 below.

identical across regions within a country (in other words, that the same variables predict participation everywhere, even though their distribution may differ across regions). The weighting procedure which we present in this paper therefore does not necessarily substitute traditional offline surveys by online surveys. Rather, it extends a small offline survey by a large online survey; cost savings may therefore primarily achieved when it comes to regional analyses.

These cost saving may seem obvious, but they are not, since after a weighting procedure has been implemented, 1000 (weighted) online observations do not correspond to 1000 observations from a random sample of the population. This fact is often ignored in the analysis of online surveys. Observations, or cells, with low selection probabilities carry large weights – in the case of internet surveys, think of older women with low education who have low rates of internet access. Similarly, those cells with high selection probabilities, such as young men with high education, carry very small weights. As a result, the effective sample size of the weighted online survey (i.e., the size of a simple random sample that would provide precision levels approximately equal to that of the more complex sample) is much smaller; we return to this issue below. While internet surveys are generally cheaper on a per-case basis than offline surveys, it is therefore not clear whether a combined offline-online design actually offers any cost saving once the reduced effective size of the weighted online sample is taken into account. This problem is obviously crucial from a practical perspective and requires a formal statistical analysis, to which we turn in the next section.

3. The sample selection problem in online surveys

In this section, we present a behavioral model of participation in online surveys, and we derive a corresponding weighting procedure that allows to correct online survey data for sample selection. The procedure we propose consists of two steps. First, it corrects for differential access to the internet; the resulting distortion in sample composition is referred to as "internet bias" in the sequel. Second, it addresses the problem of self-selection into an online sample, conditional on having access to the internet, referred to as "participation bias" in the sequel.

The formal model of participation in an internet survey forms the basis for the weighting procedure presented in section 2.2. To introduce some notion, we distinct three vectors that each collect similar variables (i.e., variables that are functionally equivalent in the selection process that generates an online sample):

x socio-demographic variables that are predetermined for each individual (such as age, gender, education, but also the region where the individual lives)

z variables that reflect an individual's attitudes (such as risk preference or constructs such as performance ethics or social ethics)

y variables that reflect an individual's responses to the public policy questions presented in the survey (such as satisfaction with the conduct of public policy or approval of some policy reform proposal)

Ultimately, the object of interest is the vector of response variables, y. The aim of the weighting procedure is to make the online data representative – i.e., to recover from the online data a density for y which is proportional to the density of y found in the population as a whole. To this end, it is useful to describe response behavior in an online survey in terms of a data generating process, that is, as conditional distributions of the various sets of variables of interest. This data generating process corresponds to a behavioral model of online survey participation.

The joint distribution of the variables of interest in the population, denoted $f(y,z,x)$, can be represented by the following data generating process that factorizes $f(y,z,x)$ in three components:

$$f(x,y,z) = f_1(x)f_2(z\mid x)f_3(y\mid z,x).$$

The distribution of the response variables of interest, $f_3(y|z,x)$, is conditional on the socio-demographic characteristics of a person and of her attitudes, z and x. The distribution of attitudes is, in turn, conditional on socio-demographic characteristics – this is captured by $f_2(z|x)$. Finally, the distribution of the pre-determined characteristics, $f_1(x)$, is unconditional.

In order to analyze the online data, one would ideally use as a benchmark the conditional distributions of variables summarized in the population data generating process just presented. This is the formal reason why weights need to be based on a representative offline survey – while $f_1(x)$ is known from official sources such as census data, and while $f_2(z|x)$ might be recovered from other surveys, $f_3(y|z,x)$ is generally not available from other sources.

It is important to note that a traditional CAPI survey that is administered, say, by a random route protocol is not necessarily representative of the population – usually, only part of the households or persons contacted is willing to answer a survey.[6] Therefore, even data from the parallel offline survey need to be weighted to ensure that they are representative of the population. These weights are derived from official census data; this is of course a standard procedure in survey research. Formally, let $f_4(y,z,x)$ be the probability of being intercepted by the offline sampling frame (e.g., the probability of being in the sample generated by following a random-route protocol). Then the conditional probability of (y,z,x) in the offline sample is

$$f\,(y,z,x\,|\,offline) = \frac{f_1(x)\,f_2(z\,|\,x)\,f_3(y\,|\,z,x)\,f_4(y,z,x)}{\sum_x\sum_z\sum_y f_1(x)\,f_2(z\,|\,x)\,f_3(y\,|\,z,x)\,f_4(y,z,x)}\,.$$

If an offline observation can be weighted by $w_4(y,z,x) = 1 / f_4(y,z,x)$, i.e., by the inverse of the selection probability, then the weighted offline sample will have the same data generating process as the population. In general, there is insufficient information to determine w_4. However, in a carefully conducted offline survey, there will be sufficient re-contact to ensure that $f_4(y,z,x)$ is essentially independent of y and z, i.e., independent of variables other than the demographic variables contained in the vector x. Then, external data on the distribution of x, say from official census statistics, can provide weights $w_4(x)$ that are a valid approximation of $w_4(y,z,x)$. This is the standard approach to weighting in survey research. For the purpose of our analysis, we treat the (weighted) data from the parallel offline study as representative of the population.

The data generating process of the online sample diverges from the density in the population along two dimensions. As discussed earlier, these dimensions are (i) internet access and (ii) the participation decision. These dimensions are reflected by two additional densities, $f_5(y,z,x)$

---

[6] In the offline survey that was conducted as part of *Perspektive Deutschland*, this proportion (i.e., the "survey reach") was about 60%, in line with levels typically achieved by professional survey firms in Germany.

and $f_6(y,z,x)$, respectively. These densities represent the probabilities of participation at each of these two stages of the participation process, stratified by the *y*, *z*, and *x* variables. Accordingly, the conditional probability distribution of online responses given internet access and participation is

$$f(y,z,x \mid online) = \frac{f_1(x) f_2(z \mid x) f_3(y \mid z,x) f_5(y,z,x) f_6(y,z,x)}{\sum_x \sum_z \sum_y f_1(x) f_2(z \mid x) f_3(y \mid z,x) f_5(y,z,x) f_6(y,z,x)}.$$

After weighting the online data with $w_5(y,z,x) = 1/\hat{f}_5(y,z,x)$ and $w_6(y,z,x) = 1/\hat{f}_6(y,z,x)$, where $\hat{f}_5$ and $\hat{f}_6$ are estimates for f_5 and f_6 respectively, the distribution of the variables of interest (y,z,x) is

$$\frac{1}{l} \cdot f_1(x) f_2(z \mid x) f_3(y \mid z,x) \cdot k(y,z,x),$$

where

$$l = \sum_x \sum_z \sum_y f_1(x) f_2(z \mid x) f_3(y \mid z,x) \cdot k(y,z,x)$$

and

$$k(y,z,x) = \frac{f_5(y,z,x)}{\hat{f}_5(y,z,x)} \cdot \frac{f_6(y,z,x)}{\hat{f}_6(y,z,x)} \approx 1.$$

Note this density is proportional to the population density – i.e., the access and participation corrected online data are representative of the population. In the next section, we describe how the weights, $w_5(y,z,x) = 1/\hat{f}_5(y,z,x)$ and $w_6(y,z,x) = 1/\hat{f}_6(y,z,x)$, can be estimated.

The two-step participation model results in a two step procedure to determine the weights. In theory, both steps of the weighting process could be performed using only demographic variables that are available from official census statistics. However, such an approach would suffer from two related problems. First, because the set of variables available from public sources is rather limited, they would most likely be relevant in both steps. But if there are no *ex ante* exclusion restrictions, the two steps would not be separately identified. It is, however, most relevant for the substantive analysis whether biases results from internet participation in general (which applies to all online surveys), or specific participation bias rooted in the content matter of the survey. Separate identification of the weights requires that the exact composition of the vectors differs (i.e., there must be exclusion restrictions). For instance, the probability of having internet access can typically be modeled sufficiently well using a small set socio-demographic and attitude variables. The participation decision itself might depend on a much richer set of variables that includes response variables as well.

The second problem of using official census statistics only is left-out variable bias. Readily available population data is commonly not sufficiently rich in terms of variables and too unspecific to address the specific participation bias in a given survey. Since the way how survey participation depends on personal characteristics and the set of personal characteristics that play a role typically vary from survey to survey, it is in general necessary to conduct the complete online survey in parallel using traditional methods and a representative sample.

In the present analysis, we have chosen to model internet access using a standard probit model with a binary dependent variable that indicates whether the individual has internet access or not. This model is estimated using data from the offline survey in which a question on internet access was contained. This approach is essentially equivalent to the well-known method of propensity score weighting (Rosenbaum and Rubin, 1983, 1984) – the propensity score is an individual's probability of having internet access, and its inverse is the weight that this individual's response receive. This approach has the advantage that it yields direct estimates of the influence of various variables on the probability of internet access that can be compared and validated with other sources. Moreover, there exists a well-established test theory that allows to determine a parsimonious, interpretable model.

The second step of the weighting process that corrects for the participation decision is less straightforward to implement. The reason is that in contrast to the first step where in the offline sample both individuals with and without internet access could be observed, the second-stage counterfactual (i.e., self-selected non-participants) are observed neither in the online nor in the offline samples. Therefore, one cannot construct a convenient indicator binary participation variable and estimate the corresponding participation probabilities directly. Rather, weights need to be constructed indirectly based on the observed values of the (y,z,x) variables in the online and offline samples. We chose to use standard iterative proportional fitting ("raking") approach at this stage. This approach (Deming and Stephan, 1940; Ireland and Kullback, 1968) ensures that conditional distributions of response variables in the offline and weighted online samples match.

4. Implementation of the weighting procedure

In this section, we describe how the two-stage weighting scheme was applied to the *Perspektive Deutschland* data. A necessary condition for the weighting to be successful is that the online sample has reached all strata of the population that one wants to investigate. A large sample size per se does not guarantee this condition. Due to the fact that internet access is higher for men, declines with age and increases with education (see table 1 above), the critical segment of the population are older women with low education.

For the substantive analysis, we restricted the sample up-front to those 160,286 of the total 169,315 participants who are over 18 years of age and have answered the question on their regional provenance.[7] The next step was to check whether this sample could be used for substantive analysis. A first indication of demographic groups that might cause problems in the weighting procedure are low cell counts in the bivariate distribution of age and gender (table 2). Despite the huge number of observations, the cell "females, 70 years and older" contains as little as 124 participants. Stratifying further by, say, education or marital status would yield very small cell counts and extremely high weights. As the situation is only little better for the woman between 60 and 69 years of age with 761 participants, we decided to restrict the substantive analysis to the age group of between 18 years and 59 years. This reduces the size of the online sample used in the weighting procedure and the subsequent substantive analysis to 151,314 observations.

Step one: correction of the internet bias

The aim of the first step of the weighting procedure is to correct for the bias that results from differential access to, and use of, the internet. We use a probit model, calibrated on the offline sample, to predict the *ex ante* probability of using the internet for every observation in the online sample.

The dependent variable of the probit model is a binary indicator of whether an individual uses the internet. As explanatory variables, we focused on socio-demographic variables that are known to influence internet access (see table 1 above) as well as psychographic variables (i.e., x and z variables, respectively). Due to the subject of the questionnaire the psychographic variables that we use reflect dimensions such as social responsibility and willingness to engage oneself. We should note that our objective is to estimate a prediction model and not a causal model. This is relevant for the interpretation of the coefficients, in particular since some explanatory variables might be endogenous.

Table 3 contains a list of the variables we used at the two stages of the weighting process. In the first-stage probit model, the variables we use are age, sex, education, employment status, regional variables indicating the federal state and university regions, two variables reflecting social responsibility, and four variables reflecting the willingness to engage oneself. In addition, we include dummy variables that capture the interaction of age (in 10-year intervals) and gender.

The number of offline observations that can be used to estimate the internet access model is 1202. This is considerably less than the total size of the offline sample (2715 observations), for two reasons. First, the offline survey was designed to be representative of the population aged 18 and older, so by focusing on the population aged 18 to 59, we loose 732 observations or 27%. Second, the offline survey was conducted in two waves (one at the beginning of the field period of the online survey, the other at the end), and due to an unfortunate design flaw, the question on internet access was not administered in the first wave of the offline survey. This results in the loss of another 719 observations. Another 62 observations were lost due to item non-response in explanatory variables.

We started by specifying a broad model that contains all explanatory variables listed in table 3 (including dummies for the 16 federal states). In order to arrive at a more parsimonious model, we restricted the broad model by eliminating non-significant explanatory variables.[8] The excluded variables include all the regional variables and some of the psychographic

[7] 2,753 participants could not be assigned locally and an additional 6,276 were younger than age 18.

[8] The validity of the exclusion restrictions implied by dropping variables was tested with the usual likelihood-ratio tests.

variables. The results of this reduced model are reported in column A of table 4. The pseudo R^2 of this model is 18 percent (which is relatively good for a cross-section model). As a final specification check, we performed out-of-sample predictions, in which model A achieved a hit ratio of slightly less than 70 percent.

Using model A, internet access probabilities can be predicted in the online sample, and the inverse of these probabilities can be used as weights that correct for differential internet access in the population. Due to item non-response, about 12% of the online participants have missing observations on one or more variables used in the prediction model. In the online data, this is the case for about 12% of individuals. Rather than discarding these observations or narrowing the prediction model even further for all individuals, we estimate a second restricted probit model for the individuals with missing observations on the explanatory variables in model A. Model B is used for those 12% of individuals, and it only employs explanatory variables that have no missing observations in the online sample. These variables are age, sex, and regional indicators. The results of this model are reported in column B of table 4. Not surprisingly, the overall fit of model B is not as good as that of model A.

With these models at hand, the weights that correct for differential internet access can finally be constructed. These weights are normalized such that their mean is 1. The standard deviation of normalized first-stage weights is 0.5. The largest weight is 5.93, and the smallest weight is 0.5.[9] While a maximum weight of about 6 is satisfactory, it is still instructive to take a closer look at the observations that carry high weights. In the group of individuals that correspond to the top percentile of weights, the marginal distributions of key demographic variables are as follows: 100% are women; 63% are aged 40-49 and the remaining 37% are in the top age bracket (50 to 59); and 100% fall in the lowest education bracket. Once again, this result confirms that certain subgroups of the population have significantly smaller internet access probabilities and therefore receive high weights in online surveys.

[9] The ratio of the largest and smallest weights is 12. According to a commonly used rule of thumb, ratios larger than 10 are problematic. We introduce a formal approach that allows to assess problems caused by large and small weights below.

Step two: correction of the participation bias

The aim of the second step of the weighting procedure is to correct the online sample (corrected for first-stage bias from differential internet access) variations in survey participation. This step combines (i) variations in the probability of being exposed to various forms of solicitation and (ii) the participation decision itself (i.e., self-selection into the sample). Due to the lack of reliable data on individuals who have internet access and have been exposed to solicitation, we combine these two parts in the second step of the weighting procedure, and we use an iterative proportional fitting (raking) algorithm to adjust marginal response distributions in the online survey to those in the offline survey.

By construction, the raking algorithm adjusts any set of marginal distributions in the online survey that is specified by the investigator to their population (offline) counterparts. A complete match of the online and offline sample could be achieved by using all variables of interest as raking variables. However, this would result in extremely high weights and instability of the analysis of subgroups. In practice, a smaller set of raking variables is therefore used. Determining what variables to use requires taking a stand on the trade-off between a perfect match of weighted online and offline responses and avoiding large weights. Before we discuss the problem of large weights and their implications in section 4, we describe the raking process, the set of raking variables that we selected, and the quality of the weighting process in the remainder of this section.

The core set of raking variables contains basic socio-demographics and psychographic variables; table 3 contains a list of these variables. Due to the social-political focus of the questionnaire we focus on psychographic variables that reflect social responsibility, the willingness to engage oneself, and risk attitude.[10] In the raking process, we used the iterative proportional fitting (IPF) sample balancing algorithm (Deming and Stephan, 1940).

When the raking process is completed, the online survey is weighted using the combination of first-stage internet access weights and second-stage participation weights. This typically

[10] Risk attitude was not used as an explanatory variable in the first-stage probit model because of a relatively large fraction of missing observations in the online sample. In contrast to the first-stage probit model, missing observations can be dealt with easily in the second-stage raking algorithm. The raking algorithm simply holds the fraction of missing observations fixed in the marginal distributions.

leads to a more dispersed distribution of weights, with larger and smaller weights in the extremes. The distribution of final online weights is skewed to the right. The mean is again normalized to 1, the standard deviation is 2.4. The minimum weight is 0.1 and the maximum weight is 126.7. The top percentile of the distribution consists of weights larger than 11.x. Table 5 reports the demographic characteristics of these large-weight observations and reports the corresponding statistics for the German population. It can be seen that observations which receive a high weight are predominantly female, elderly, and are poorly educated. Note that there is no regional (East vs. West) effect on the incidence of large weights.

In order to avoid instability in case of detailed analyses, we cap the weight distribution roughly at the top percentile by setting the 1123 largest raw weights to 12. This approach is heuristic; more formal approaches are discussed in Kalton and Flores-Cervantes (2003).

Comparison of weighted online and offline responses

The success of the weighting procedure can be evaluated by a comparison of the response distribution in weighted online sample with that in the representative offline sample.

The substantive analysis of the survey concerned 76 variables. Most of the questions corresponding to these variables offered a six-point rating scales as the response format. For ease of presentation, the substantive analysis of *Perspektive Deutschland* focused on a specific summary statistic, namely, the fraction of participants who reported either the top two or the bottom two values, depending on the question. For instance, the substantive analysis would be based on statements such as: “65% of respondents said that they are ‘very satisfied’ or ‘satisfied’ with the quality of life in the town or region where they live.”

In table 6, we report the absolute value of the difference between these “satisfaction fractions” in the weighted online and in the offline sample for a subset of the 76 variables. Specifically, we report statistics for five variables that have not been part of the weighting procedure. These variables were those that were most prominently discuss in the media. The largest absolute deviation between the proportion of affirmative responses in the offline and unweighted online sample is 12 percentage points. In all five cases reported in the table, the differences between offline and online samples shrink after weighting. In one case, the difference is still substantial (and statistically significant at conventional confidence levels) – about seven percentage points. Similar findings hold for the whole set of 76 response variables: Percentage point deviations are rather small after weighting, with the majority being smaller than five percentage points. The finding that well-calibrated weighting schemes work well for most, but not for all variables confirms findings in the existing literature on online surveys, such as Schonlau et a. (2004).

5. Consequences of weighting for sub-sample analysis

One of the potential advantages of large samples in internet surveys such as *Perspektive Deutschland* is that they allow a much more detailed analysis at the level of sub-samples such as regions. This is possible under the assumption is that there are no differences in internet access or the decision to participate in the internet survey cross the sub-samples in questions (such as regions). Under this assumption, a selection correction implemented using an offline sample that is representative of the national population will also be valid for each sub-sample, and the corresponding online weights can be used for sub-sample analyses as well.

In the case of *Perspektive Deutschland*, we are not aware of any convincing *a priori* reasons that would cast doubt the validity of this maintained assumption. However, sub-sample analysis might encounter practical problems since weighting might reduce effective sample sizes dramatically once the analysis looks at regional sub-samples. In the remainder of this section, we explore the concept of effective sample sizes in more detail. This measure allows to assess whether a given sub-sample of the weighted online survey can yield valid results.

It is well known that the variance of a sample statistic such as the sample mean increases when individual observations are weighted; for a clear exposition, see Deaton (1997, pp. 44–47). The variance of a statistic increases in the dispersion of weights. In particular, having larger weights in the sample results in larger standard errors and confidence bands. This can be seen formally from the formula for the variance of the weighted sample mean of some variable y,

$$\hat{V}(\bar{y}) = \frac{N}{N-1}\sum_{i=1}^{N}\left(\frac{w_i}{\sum_{k=1}^{N} w_k}\right)^2 (y_i - \bar{y}).$$

Here, N is the sample size, and w_i is the weight attached to observation i with variable y_i. Since weights appear squared in the numerator, large weights increase the variance. From this formula, the effective sample size, N_{eff} can be derived[11]:

$$N_{eff} = \frac{\left(\frac{1}{N}\sum_{i=1}^{N} g_i\right)^2}{\frac{1}{N}\left(\sum_{i=1}^{N} g_i^2\right)} \cdot N =: E \cdot N \ ,$$

where E is referred as the efficiency of the weighted sample. As the average weight of the online-survey is 1 by definition, we just have to calculate the average of the squared weights: In the weighted online survey, with weights being capped at 12, as described above, this number is 4.22. The effectiveness of the sample is 1/4.22=0.24. Without capping, this number would drop dramatically to 0.09 (even though only about one percent of observations have a weight larger than 12). In traditional offline surveys administered by random route protocols, the effectiveness of the weighted sample is approximately 0.6 when weights are constructed from census data. It is important to note that due to the incidence of more extreme weights, online surveys are less effective than traditional surveys, and the plain number of participants may be misleading even if it is very large. In the case of *Perspektive Deutschland*, we obtain an effective sample size of $N_{eff} = 0.24 \cdot 151{,}347 = 35{,}876$ participants.

When it comes to analyzing sub-samples, the concept of the effective sample size is helpful to determine whether valid analyses are feasible (without looking at the variance of weighted means for specific variables to see whether a pair-wise test would have any power). One simply needs to compute the effective sample size for each sub-sample of interest. Sample effectiveness varies across sub-samples to the extent that the distribution of weights varies

[11] In survey research, the "effective sample size" is the size of a simple random sample that would provide approximately equal precision levels to that of the more complex sample.

across sub-samples. For instance, outliers might have significant influence in sub-samples even though they do not matter much for the full sample.

In table 7, we report effective sample sizes for the 10 smallest and 10 largest of the 97 regions (*Raumordnungsregionen*) considered in the substantive analysis of Perspektive Deutschland. We also report the fraction of respondents who reported that they are very satisfied with the quality of live in the town or region where they live, together with two standard errors – one computed not taking the online sample weights into account, the other computed according to correct formula reported above. As one can see, weighting increases standard errors and confidence intervals generally. While an increase from a 1 percent standard error to a 2 percent standard error (as in the large regions such as Berlin) will not affect substantive analyses, standard errors may increase dramatically in smaller regions. In the case of the Altmark, the increase in the standard error is 7 percentage points. This effect is exacerbated since standard error for fractions are larger for small or large fractions, that is, in regions with high or low average levels of satisfaction; in that sense, the Altmark with its combination of low satisfaction and small number of observations is the most critical sub-sample in *Perspektive Deutschland*. Also, note that without capping of large weights, effective sample sizes in regional sub-samples would be even smaller, and small regions such as the Altmark could not be analyzed any more.

Measures such as the effectiveness of a weighted sub-sample can serve as helpful warning signs for sub-sample that might be problematic. In the end, substantive conclusions might still be warranted, depending on the question at hand and weighted standard errors. For instance, in the case of the Altmark, even the large standard error of 11 percentage points does not change the conclusion that this region is in the group of region with the lowest levels of regional satisfaction in the sample.

6. Conclusions

In this paper, we presented a two-step procedure to correct sample selection problems in online surveys. This procedure is based on a conceptual model that splits selection into an online sample into two components: access to the internet and self-selection into the survey, conditional on internet access. The two-step procedure derives weights that reflect the variables that drive these two aspects of the sample selection process. We showed how this procedure was implemented to derive weights for a large-scale online survey with a total of about 170,000 participants and a parallel offline survey administered using a traditional CAPI interview and a random-route protocol.

The case study highlights the practical relevance of some well-known problems of internet surveys. Some groups of the population have very low rates of internet access, in particular, older women with low education. In practice, this problem forced us to restrict the substantive analysis to individuals younger than age 60. Even in this restricted sample, a small fraction of weights was quite large. We used a heuristic approach, essentially capping the top percentile of the distribution of weights.

The present paper was intended as a case study of the strengths and weaknesses of large-scale online surveys. The methods we used are sometimes heuristic – a choice we deliberately made because our approach should be simple enough for everyday use in marketing and opinion research practice. Our analysis could be extended in two directions. First, a more formal analysis of the properties of the weights generated by our two-step weighting procedure could be conducted along the lines of Lu and Gelman (2003). Second, as we mentioned above, the behavioral model of internet survey participation that forms the basis of our two-step weighting procedure could be combined with a model of the response process itself. This joint model could then be estimated with more efficient methods for choice-based samples following Cosslett (1981) and Hsieh et al. (1985). Such an approach was used, among others, by McFadden et al. (1985) and Hellerstein and Imbens (1999), but has, to our knowledge, not yet been applied to large-scale internet survey. Both extensions are beyond the scope of this case study but could be fruitfully developed in future research on participation and response behavior in internet surveys.

For marketing and opinion research practice, one of the main questions when it comes to online surveys is – do they really make sense? It is often argued that online surveys are cheaper than offline surveys on a per-case basis. As we have seen in this paper, once one takes into account that the effective sample size of a weighted online survey is only about 25% of the number of participants. From this perspective, an online survey will break even once its per-case cost is less than 25% of the per-case cost of a traditional survey.

Another important aspect is that in order to compute online weights, data that are representative of the target population are required. As we have argued in this paper, the process of self selection into an online sample is complex. Weights that are based only on socio-demographic variables (which can be obtained from official census statistics) are unlikely to be sufficient to derive valid weights. The main reason for this is that at the second stage of the self-selection process – when individuals with internet access decide whether they should participate or not – psychographic variables will play a major role. One option is to use data on the distribution of such variables in the population that can be obtained from other publicly available surveys, although this might restrict the choice variables. Alternatively, a parallel offline survey can be used such as in the case of *Perspektive Deutschland*. A parallel offline survey has the advantage that the quality of the weighting process can be evaluated directly, but it is of course costly.

In marketing and option research practice, the most promising approach seems to us to conduct traditional surveys with representative samples from time to time (say, at an annual or bi-annual frequency) and to use online surveys at higher frequencies to obtain up-to-date data on topical issues. Weights for these high-frequency internet surveys could be based on the low-frequency offline survey. This approach offers significant cost savings since the representative sample has to be interviewed at a much lower frequency than the internet sample. Internet surveys can then show their full potential, namely speed, flexibility, cost effectiveness, and innovative design options, relative to traditional modes such as telephone interviews.

Literature

Bosnjak, M., T. L. Tuten, and W. Bandilla (2001): Participation in web surveys: A typology. *ZUMA-Nachrichten*, 48, 7–17.

Clayton, L. and S. Werking (1998): Business surveys of the future: The World Wide Web as a data collection methodology. In M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nichols II, and J. M. O'Reilly (Eds.), *Computer Assisted Survey Information Collection*, New York, NY: 543–562.

Cosslett, S. R. (1981): Maximum likelihood estimator for choice-based samples. *Econometrica*, 49, 1289–1316.

Couper, M. P. (2000): Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64, 464–494.

Deming, W. E. and F. F. Stephan (1940): On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Journal of the American Statistical Association*, 35, 615–630.

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 Economic and Social Measurement*, 5, 475–492.

Hellerstein, J. K. and G. W. Imbens (1999): Imposing moment restrictions from auxiliary data by weighting. *Review of Economics and Statistics*, 81, 1–14.

Hsieh, D. S., C. F. Manski, and D. McFadden (1985): Estimation of response from augmented retrospective observations. *Journal of the American Statistical Association*, 80, 651–662.

Ireland, C. T. and S. Kullback (1968): Contingency tables with given marginals. *Biometrika*, 55, 179–188.

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

Lu, H. and A. Gelman (2003): A method for estimating design-based sampling variances for surveys with weighting, poststratification, and raking. *Journal of Official Statistics*, 19, 133–151.

Manski, C. F. and S. R. Lerman (1977): The estimation of choice probabilities from choice based samples. *Econometrica*, 45, 1977–1988.

McCullough, D. (1998): Market research on the web. *Communication World*, 15, 29–31.

McFadden, D. (2001): Observational studies: Outcome-based sampling. In N. J. Smelser and P. B. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences*, 10804–10808. Amsterdam: Elsevier.

McFadden, D., S. Cosslett, G. Duguay, and W. S. Jung (1977): *Demographic data for policy analysis*. Urban Travel Demand Forecasting Project, Final Report, Volume VIII, Institute of Transportation Studies, University of California, Berkeley.

McFadden, D., C. Winston, and A. Börsch-Supan (1985): Joint estimation of transportation decisions under nonrandom sampling. In A. Daughety (Ed.), *Analytical Studies in Transport Economics*, Cambridge, UK: Cambridge University Press, 137–157.

Rosenbaum, P. R. and D. B. Rubin (1983): The central role of the propensity score in observational studies of causal effects. *Biometrika*, 70, 41–55.

Rosenbaum, P. R. and D. B. Rubin (1984): Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79, 516–524.

Schonlau, M., R. D. Fricker, Jr., and M. N. Elliott (2002): *Conducting Research Surveys via E-mail and the Web*. Santa Monica: RAND.

Schonlau, M., K. Zapert, L. Payne Simon, K. Sanstad, S. Marcus, J. Adams, M. Spranca, H.-J. Kan, R. Turner, and S. Berry (2004): A comparison between a propensity weighted web survey and an identical RDD survey. *Social Science Computer Review*, 22, 128–138.

Taylor, H., J. Bremer, C. Overmeyer, J. W. Siegel, G. Terhanian (2001): The record of internet-based opinion polls in predicting the results of 72 races in the November 2000 U.S. elections. *International Journal of Market Research*, 43, 127–135.

Table 1: Internet use: Germany vs. United States

		Fraction of internet users in the adult population	
		Germany	U.S.
Overall		36.4	54.7
Gender	Female	31.5	55.0
	Male	41.4	54.3
Age	18-29	59.6	65.7
	30-39	48.2	66.1
	40-49	41.9	63.6
	50-59	32.0	57.0
	60-69	13.2	34.9
	70 +	3.3	15.0
Education	Low education	25.0	34.3
	High education	69.0	61.4
Region	West Germany	36.9	N/A
	East German	31.4	N/A
	Berlin	46.4	N/A

Notes: For Germany, the "high education" category contains all degrees which entitle at least to study at a polytechnic or other higher education institution. For the U.S., "high education" is a high-school diploma or more.
Sources: Germany: Offline sample of *Perspektive Deutschland*, Winter 2001. U.S.: Current Population Survey (CPS), September 2001.

Table 2: Bivariate distribution of age and gender in the online sample (age 18 and older) and in the population

Age group	**Female**			**Male**			**Total**		
	Online Counts	Online Percent	Pop. Percent	Online Counts	Online Percent	Pop. Percent	Online Counts	Online Percent	Pop. Percent
18-29	19698	46.2%	16.2%	41205	35.0%	18.2%	60903	38.0%	17.2%
30-39	11809	27.7%	19.5%	31703	27.0%	22.3%	43512	27.1%	20.9%
40-49	6972	16.3%	17.2%	21828	18.6%	19.1%	28800	18.0%	18.1%
50-59	3294	7.7%	14.2%	14805	12.6%	15.3%	18099	11.3%	14.8%
60-69	761	1.8%	14.8%	6560	5.6%	14.8%	7321	4.6%	14.8%
70+	124	0.3%	18.0%	1527	1.3%	10.3%	1651	1.0%	14.3%
Total	42658	100%	100%	117628	100%	100%	160286	100%	100%
Online Percent	26.6%			73.4%			100%	100%	
Pop. Percent	51.8%			48.2%			100%		

Notes: Population data are from the German *Mikrozensus*.

Table 3: Variables used in the two-step weighting procedure

Variable	Source	First stage (probit model)	Second stage (raking)
Interaction of gender and age (brackets of 10 years)	*Mikrozensus*	YES	YES
Education (5 categories)	*Mikrozensus*	YES	YES
I consider it important to achieve more than others.	Offline		YES
You do not always have to think about getting ahead; sometimes you should be content with what you have.	Offline	YES	NO
Employment (2 categories)	*Mikrozensus*	YES	NO
Employment (8 categories)	*Mikrozensus*	NO	YES
I turn my goals and ideas of achievement into reality	Offline	YES	NO
I want to achieve in line with my abilities	Offline	YES	YES
I would like to rise in society	Offline	YES	YES
I feel responsible for society	Offline	YES	YES
I see society as a form of insurance: when I am doing well I contribute, and when I am doing badly I get something back	Offline	YES	NO
Regional indicator (16 categories one for each federal state)	Offline	YES	NO
Regional indicator (97 regions)	*Mikrozensus*	NO	YES
University region (2 categories)	*Mikrozensus*	YES	YES
Income (5 categories)	*Mikrozensus*	NO	YES
Interaction of income and age categories (16 categories)	*Mikrozensus*	NO	YES
Children (2 categories)	Offline	NO	YES
I am prepared to take risks in my professional career	Offline	NO	YES
I am prepared to take risks in financial investments	Offline	NO	YES
Intention to start ones own business (4 categories)	Offline	NO	YES

Table 4: Determinants of internet access: probit estimates

	Coefficient (asymptotic p-value)	
Variable	**Model A**	**Model B**
Age 18-20, female (D)	0.400	-0.097
	(0.327)	(0.782)
Age 20-29, female (D)	0.161	0.279
	(0.366)	(0.121)
Age 40-49, female (D)	-0.496	-0.383
	(0.002)***	(0.009)***
Age 50-59, female (D)	-0.598	-0.655
	(0.001)***	(0.000)***
Age 18-20, male (D)	0.955	0.372
	(0.023)**	(0.270)
Age 20-29, male (D)	0.402	0.596
	(0.037)**	(0.002)***
Age 30-39, male (D)	0.077	0.272
	(0.638)	(0.069)*
Age 40-49, male (D)	0.065	0.118
	(0.695)	(0.451)
Age 50-59, male (D)	-0.247	-0.067
	(0.169)	(0.692)
Education; high school student (D)	-0.429	
	(0.330)	
Education; entitled to study at a univ. (D)	1,211	
	(0.000)***	
Education; university degree (D)	1,285	
	(0.000)***	
Employment status (D)	-0.278	
	(0.012)**	
I feel responsible for society. (D)	-0.150	
	(0.021)**	
You do not always have to think about getting ahead; sometimes you should be content with what you have. (D)	0.152	
	(0.040)**	
University region (D)		0.103
		(0.243)
Constant	-0.218	-0.158
	(0.094)*	(0.148)
Log Likelihood	-674	-781
Hit ratio	69%	60%
Pseudo-R^2	.18	.06
Observations	1202	1202

Notes: Significance levels: *, 10%; **, 5%; ***, 1%.

Table 5: Descriptive statistics of online observations with large weights

		Percentage of Online observations with weights larger than 12	**German population aged 18-59**
N		1123	
Gender	Female	81.3%	50.9%
	Male	18.7%	49.1%
Age	18-19	0.2%	4.0%
	20-29	7.9%	20.2%
	30-39	25.0%	29.4%
	40-49	29.7%	25.6%
	50-59	37.2%	20.8%
Region	West Germany	77.1%	77.3%
	East German	19.6%	18.7%
	Berlin	3.3%	4.0%
Education	High school student	0.1%	2.3%
	Low education	98.9%	70.5%
	High education	1.0%	27.2%

Notes: The “high education” category contains all degrees which entitle at least to study at a polytechnic or other higer education institution. Data on gender and age are from the German *Mikrozensus*. Data on region and education are from the Offline-Sample.

Table 6: Alignment of selected response variables in the weighted online and offline samples

Statement	Proportion of affirmative responses			Difference of means between offline and …		Number of observations	
	Offline	Online unweighted	Online weighted	Online unweighted	Online unweighted	Offline	Online
Taking it all in all, you can have a very good life in a country like Germany.	0. 66	0.72	0.66	0.05	-0.01	1974	151314
Taking it all in all, you can have a very good life in the city or region I live in..	0.69	0.70	0.65	0.01	-0.04	1972	151314
I would like to have more say in the way I carry out my tasks at work, even if I were ultimately held responsible for the results.	0.54	0.64	0.60	0.10	0.07	1282	58340
Would you welcome a payment scheme that is based more on your individual performance?	0.41	0.53	0.43	0.12	0.03	1276	58326
Do you think that increased private provision will be needed to supplement public social security in the future?	0.43	0.52	0.41	0.09	-0.02	1935	80601

Table 7: Effective sample sizes and confidence intervals for regional sub-samples

Rank by region size	Region	Fraction of "I agree" responses (in percent)	"Naïve" calculation		Efficiency without cap			Efficiency with cap at 12		
			Raw cell count	Half-width of 95% confidence interval	Efficiency factor	Effective cell count	Half-width of 95% confidence interval	Efficiency factor	Effective cell count	Half-width of 95% confidence interval
1	Berlin	55	6246	1	0,13	1269	3	0,20	1888	2
2	Düsseldorf	68	5478	1	0,18	1143	3	0,24	1574	2
3	Rhein-Main	72	4968	1	0,15	1129	3	0,20	1481	3
4	Stuttgart	77	4839	1	0,14	855	3	0,22	1374	3
5	München	78	4405	1	0,13	1051	3	0,18	1484	3
6	Duisburg/Essen	65	4239	1	0,18	573	4	0,27	877	3
7	Köln	75	3956	1	0,09	403	5	0,23	1064	3
8	Hamburg	75	3185	2	0,14	637	4	0,22	992	3
9	Bielefeld	74	3018	2	0,14	323	5	0,27	647	4
10	Münster	75	2885	2	0,12	302	6	0,24	601	4
...	...	...	...	...	...	...	...	...	...	...
88	Siegen	68	754	3	0,06	36	16	0,24	150	8
89	Bayerischer Untermain	82	754	3	0,22	149	8	0,28	183	7
90	Mecklenburgische Seenplatte	32	606	4	0,21	66	12	0,33	101	10
91	Lüneburg	72	604	4	0,13	68	12	0,26	135	8
92	Uckermark-Barnim	35	604	4	0,20	86	11	0,32	139	8
93	Osthessen	62	604	4	0,23	102	10	0,30	132	9
94	Bremerhaven Umland	64	604	4	0,24	83	11	0,39	135	8
95	Südheide	68	603	4	0,22	90	10	0,30	122	9
96	Schleswig-Holstein Süd-West	77	455	4	0,27	87	11	0,30	96	10
97	Altmark	28	453	4	0,18	37	16	0,32	65	12
Total (Germany)		65	151314	0	0,09	13539	1	0,24	35876	0

Note: The target question is "All in all, the quality of life in the town or in the region, where I live, is very high."