Deviations from the population and optimal weights

Häder, Sabine; Gabler, Siegfried

CHAPTER 3

DEVIATIONS FROM THE POPULATION AND OPTIMAL WEIGHTS

SABINE HÄDER AND SIEGFRIED GABLER

3.1 Introduction

Nonresponse has become an important problem in the empirical social sciences.[3] This problem plays a role in face to face studies as well as in telephone and mail surveys. Presently, the percentage of nonresponse frequently ranges from 30-50%.[4] These losses due to nonresponse may lead to systematic biases in the samples which result in biased estimates. The application of weighting procedures is a usual way to compensate for this bias.

The weighting process changes the relative importance of the respondents. Technically speaking, weighting means the attachment of numbers to elements of the responding population. As a consequence, after weighting the sample profile of a variable of interest should be more similar to the population profile than without weighting. But for this logic to apply, it should be taken for granted that within each weighting class, the profiles of the survey variables are very similar for the responding and nonresponding parts of the sample (Elliot, 1991: 5).[5] In this way it is possible "...to make the sample data we collect more representative of some population data we are trying to measure or estimate" (INRA, 1994: 26).

In general, the adaptation cannot be achieved for all variables at the same time. Therefore, some variables for which the population profiles (e.g. gender, age, household size) are known are usually selected to act as so-called active variables. Testing the effect of weighting is possible only for a few additional variables not used for weighting for which we also know the distributions from official statistics, like income, marital or professional status. But the key assumption for successful weighting is that the bias of the other (passive) variables like attitudes or behaviours is reduced by the weighting procedure as well. However, resulting

3 For a classification of nonresponse see Kish, 1965: 532-534.

4 Some examples to illustrate this statement: The ALLBUS 1994 - a face to face study in Germany - had a nonresponse rate of 46 percent (Koch et al., 1994: 82). In the National Readership Survey (UK) a nonresponse rate of 39 percent was stated. Also, in American surveys the nonresponse rates reach about 40 percent nowadays (Bradburn, 1992: 392). The average rate only of refusals and hangups of the FORSA European Telephone Survey 1994 was about 40 percent (FORSA, 1994: 7-13) of the Net Sampe Pool.

5 Of course this condition cannot be tested from the survey data. So we have to assume that the nonrespondents are a random subsample in each weighting class.

changes in passive variables due to weighting cannot be controlled by comparisons with the distributions of the total population because we do not have population profiles from sources other than from empirical surveys which may be biased themselves.

Because of these difficulties the weighting procedure is discussed controversially by the empirical social scientists (for a description of different points of view in the matter see Gabler et al., 1994).

By showing the impact of weighting factors in the concrete framework of the Eurobarometer Experiment 1994, we want to contribute to this discussion. First of all, it will be tested how the samples represent the population in a comparison of different national studies. Therefore, we will show the results of a comparison of the profiles of some selected variables from the national samples and the populations. Following this, we will describe the statistical characteristics of the weights we have used. Finally, we will pursue the question whether there are differences in the distributions of the passive variables before and after weighting.

As a result of our analysis it should therefore be possible to show typical patterns of the impact of weighting in the framework of these European studies.

Table 3.1 Sample sizes of selected studies of the Eurobarometer Experiment

Nation	EB41.0	EB41.Panel	FORSA
Belgium	1087	234	501
France	1034	341	500
Spain	1003	731	500
East Germany	1058	-	500
West Germany	1064	-	500

For reasons of parsimony and clarity, we shall select only a few countries for our investigation: Belgium, France and Spain have been chosen because they are also used in the panel study.[6] We have included West Germany and East Germany in our investigation because of the large differences between these two parts of Germany concerning the availability of telephones in private households. In 1994, more than 9 out of 10 households in West Germany had a telephone, while the level of telephone availability in private households had reached only about 50% in East Germany till then (Drews, 1994; Häder, 1994). This low level of telephone penetration in East Germany contains a great risk of a major sampling error because not all of the households of the target population have a known positive chance of being included in the sample. Besides the nonresponse error, this fact can also negatively affect the accuracy of the estimates. Therefore, the differences in the representation of the target population between the face to face and the telephone survey in East Germany should be particularly observed.

6 We do not analyse the samples of the panel studies because the number of realized valid cases is too small for our investigation.

3.2 Comparison of the distributions of selected active variables of the national samples with statistical data on the target populations

A comparison of the distributions of selected demographic variables (age, gender, household size) from the different samples with the reference data based on the national statistical yearbooks shows how large the differences in all variables and nations are (see Table 3.2, 3.3 and 3.4).

Table 3.2 Percentages of males in different nations and studies as well as in the target populations

Nation	EB41.0	FORSA	Reference data
Belgium	49.0	48.0	48.3
France	48.6	39.9	48.1
Spain	48.8	37.8	48.4
East Germany	49.4	44.2	47.4
West Germany	51.0	42.6	48.1

Table 3.3 χ^2 - statistics for the distributions of age groups[7] in EB41.0, in FORSA and reference data based on the statistical yearbooks

Nation	χ^2- Statistics for Eurobarometer-Reference	χ^2- Statistics for Telephone-Reference
Belgium	1.27	4.39
France	9.61	1.34
Spain	1.49	19.63
East Germany	10.90	19.57
West Germany	11.11	7.08

Table 3.4 χ^2 - statistics for the distributions of household sizes[8] in EB41.0, in FORSA and reference data based on the statistical yearbooks

Nation	χ^2 - Statistics for Eurobarometer-Reference	χ^2 -Statistics for Telephone-Reference
Belgium	56.35	41.95
France	57.05	38.33
Spain	40.25	29.13
East Germany	51.82	51.87
West Germany	45.63	70.68

7 For the age groups we chose the following categories: 1.: 15-29 years; 2.: 30-44 years; 3.: 45-59 years; 4.: 60 years and older.

8 For the household size we chose the following categories: 1.: one person; 2.: two persons; 3.: three persons; 4.: four and more persons.

In the two Tables 3.3 and 3.4 we show the χ^2 - statistics for the test on similarity of the expected and observed marginal distributions for the variables age and household size. The critical value for an α-level of 0.05 and 3 degrees of freedom is 7.81.

For all countries - in part large (household size) - differences between the sample distributions and those of the target populations (adults of 15 years and older) have to be stated as the computed statistics show. Altogether in 15 out of 20 tested cases the differences between the distributions are significant. This means that all samples are biased demographically. Therefore, biases in other variables due to nonresponse or (other) sampling errors can be expected as well. The usual way in social and market research to deal with this problem is - as stated above -

1. to assume a fairly high correlation between demographic and all other survey variables, and
2. therefore, to adjust some of the demographic variables to their known distributions of the target population.

In the following we want to pursue this approach and - after that - discuss the results we have obtained with this method.

3.3 Effects of weighting

Before presenting our special weighting procedure we will offer some remarks on weighting in general. We distinguish two kinds of weighting:

a) *Design weights* (also called preweights): Sometimes the sampling design requires weighting of the observations to avoid bias. If the probabilities of selecting the units are not equal, we should weight the units with the inverse of the inclusion probabilities. This leads to the well-known Horvitz-Thompson estimator which is unbiased for the population total (or mean). This way of weighting is useful provided that the sample design has been realized precisely and the inclusion probabilities are exactly known.

However, analyses of the fieldwork of face to face studies as well as telephone surveys have shown that in reality this is usually not the case. Also, in the studies of the Eurobarometer Experiment considerable rates of refusals existed. For example, for the telephone surveys among the five countries discussed in this contribution refusal rates range from 19.6% of the Net Sample Pool in East Germany to 49.9% in Belgium (FORSA, 1994: 7, 13).

In several investigations it was found that there is no guarantee for improving the estimator by using design weights.[9] Anyway, as a practical problem for our investigation of the Eurobarometer Experiment we have to state that the design weights to correct for the unequal chance for households of different sizes to be selected, cannot be determined because the necessary information for the calculation is not available from the FORSA telephone surveys. As a consequence, we will do our analyses without this mode of weighting.

9 Rothe stated as a result of his investigations, „daß bei personenbezogenen Variablen die Hochrechnung ohne Verwendung einer Gewichtung eher bessere Werte liefert als die theoretisch korrekte Gewichtung mit der reduzierten Haushaltsgröße“. (see Rothe, 1994: 71)

b) *Cell weights* (also called postweights): After collecting the data, the sample profiles of some characteristics, i.e. gender or age, may differ from the corresponding population profiles which are known from external sources. Poststratification is the usual way to include this information into the estimation. If the multivariate distribution for these characteristics is unknown but some marginal distributions are known, we can try to find weights which yield representative estimators in the sense of (Hájek, 1981). This means that after weighting the sample profiles and the population profiles are identical for each of these characteristics. The solution of this problem is not unique. Usually the number of solutions is infinite. A requirement with regard to the weights should be that they are as close as possible to one. One popular solution is the **I**terative **P**roportional **F**itting solution also known as raking. The algorithm has its origin in the paper of Deming and Stephan (1940).

The weights can be found by minimizing the objective function (similar to the discrimination information or Kullback-Leibler-information) which represents the distance between the weights w_i and 1, the unweighted case,

$$\sum m_i (w_i \ln(w_i) - w_i + 1)$$

where m_i is the number of elements in cell i and w_i is the weight which we attach to all units of cell i. The constraints under which the objective function has to be minimized are that some marginal distributions of the weighted units are representative. These constraints can be represented by a matrix equation of the form $A\underline{m}_w = \underline{n}$ with known restriction matrix A , the vector $\underline{m}_w=(m_1w_1, m_2w_2, ...)'$ and the vector $\underline{n}=(n_1, n_2, ...)'$ of marginal cell frequencies.

Raking estimates are not maximum likelihood estimates of the cell proportions when the observed data are a random sample from the target population, but they are consistent and best asymptotically normal as Ireland and Kullback (1968) show.

If we do not choose the logarithmic function as an objective function which we want to minimize but the quadratic function (**L**east **S**quares, minimum variance criterion)

$$\sum m_i (w_i - 1)^2$$

we get weights with the highest efficiency[10]. The solution is given by

$$\underline{m}_{WL} = A(AA')^{-1}\underline{n}$$

which is the Moore-Penrose-Inverse of $\underline{n}$. It may be that the solution in the quadratic case is no longer nonnegative. Usually, weights less than a positive minimal number g_0 or higher than g_1 are truncated to g_0 or g_1, respectively.[11]

3.4 Restrictions for the construction of the weighting factors in the framework of the Eurobarometer Experiment

According to our research interest we wanted to construct weights in such a way that in all countries, for both the face to face and the telephone survey, an adjustment to the same variables is performed. By doing this we would be able to compare the statistical characteristics of the weights for the two samples in each country. Furthermore, it would be useful to analyse the impact of the weighting procedures on the profiles of selected passive variables.

First of all we had to solve some practical problems concerning the active variables:

- The telephone survey done by FORSA contains only 500 cases per nation. For reasons of consistency empty cells should be avoided in a weighting procedure. Thus we had to decide whether we wanted to use fewer, but more finely subdivided active variables or broader categories for more variables.
- In the questionnaire of the telephone survey the demographic questions are only partly replicated from the Eurobarometer study. For example, questions on the marital status of the respondents and on the number of children living in the household are missing. Besides, in some other eligible variables different categories have been used in the questions of the two surveys we want to compare. These facts have reduced the number of possible active variables for our analysis.
- The basis for reference statistics were the annual Statistical Yearbooks of the different countries and the German Statistical Yearbook 1994 for Foreign Countries (Statistisches Bundesamt, 1994), respectively. The tables in the different National Yearbooks are not standardized but vary from country to country. This led to problems in finding comparable distributions of eligible active variables. Therefore, we decided to take the reference data for the joint distributions of age and gender out of the statistical yearbooks of the different countries. As data source for the marginal distribution of the household sizes we used the Statistical Yearbook 1994 for Foreign Countries where the presentation was the same for all five countries.

---

10 For the definition of efficiency see chapter 3.6.2.

11 An overview of the construction of weights can be found in Alexander (1987), Deville et al. (1992, 1993), Gabler (1994), Zaslavsky (1988). A comparison of the various solutions is given by Little and Wu (1991).

3.5 Realization of the weighting procedure

Considering the problems mentioned above, we selected the following active variables to be adapted in the weighting procedure:

Age (4 groups) * Gender (2 groups) * Household size (4 groups) = 32 cells for France, Belgium, Spain, East Germany and West Germany (both samples).

As we have shown in the Tables 3.2, 3.3 and 3.4, the distributions of these variables in the national samples (face to face study and telephone survey) differ considerably from those of the reference data. Therefore, an adjustment of these selected variables seemed to be useful.

Moreover, the selection of age, gender and household size as active variables is frequently done in social as well as in market research[12]. Sometimes weighting by these factors is interpreted as a cure for biased samples. Therefore, there is a need to observe the impact of weighting on the passive variables carefully. Besides we wanted to compensate for the missing of design weights by including the household size into the group of active variables.

Because we had no joint distribution for the variables age, gender and household size for each nation, which would be the condition for simple cell weighting, we used the two different ways as already referred to for determining the weighting factors. We therefore obtain two solutions, the IPF-Solution and the LS-Solution.

3.6 Results of the analysis of the weighting factors

3.6.1 Ranges of the weighting factors

In the following we want to describe the factors we have obtained as a result of the two routines to be used for the weighting procedures. In Table 3.5 and Table 3.6 we present the ranges of the weighting factors for each country.

Table 3.5 Ranges of the cell weights for EB41.0 and for FORSA for the selected nations (IPF-Solution)

Nation	EB41.0	FORSA
Belgium	0.580 - 1.919	0.472 - 2.276
France	0.637 - 1.901	0.381 - 2.159
Spain	0.696 - 2.166	0.939 - 3.006
East Germany	0.579 - 1.901	0.268 - 5.492
West Germany	0.770 - 1.618	0.475 - 2.901

12 These active variables are for instance an important part of the weighting procedure for the „Media-Analyse“ which is well known in market research as a highly reliable study (see Rothe and Wiedenbeck, 1994).

Table 3.6 Ranges of the cell weights for EB41.0 and for FORSA for the selected nations (LS-Solution)

Nation	EB41.0	FORSA
Belgium	0.466 - 1.767	0.253 - 2.018
France	0.599 - 1.741	0.085 - 1.873
Spain	0.639 - 1.995	0.289 - 2.455
East Germany	0.506 - 1.762	- 0.453 - 3.390
West Germany	0.758 - 1.579	0.300 - 2.409

For both solutions, in all nations the ranges of the weighting factors for the telephone survey are larger than the ranges of the factors for the Eurobarometer study. In particular, it should be noted that some negative weighting factors for the telephone survey in East Germany resulted from the LS-Solution.

As previously mentioned, it is a requirement that the weights should be as close to one as possible. The Eurobarometer study is closer to the achievement of this goal in all nations. Furthermore, the sizes of the weights for the telephone study are mostly larger. So we have a first indication concerning the comparison of the quality of the two samples.

3.6.2 The efficiency of weighting procedures

"A useful measure of the effect of unequal probability sampling on precision is provided by the 'effective sample size' or ESS. It measures the size of an equal probability sample that would produce the same precision as the unequal probability design under consideration." (Elliot, 1991: 8)[13].

The efficiency (i.e. ESS as a proportion of the actual sample size) shows what proportion of the original sample size an unweighted random sample with the same variance as the weighted random sample would have. Efficiency also can be used for the comparison of the quality of two samples. In that case different efficiencies mean that the samples represent the population with different precision.[14]

In this sense we want to interpret the efficiencies in Table 3.7 and 3.8 for the Eurobarometer study and the telephone survey for our selected nations and both weighting solutions as control criteria for the quality of the different samples.

13 "The main assumption under which the formula is derived is that the true population variances are equal in the groups having different weights, although it also assumes independent simple random sampling in the different groups. In many situations theses assumptions may be reasonable enough." (Elliot, 1991: 8)

14 "Man kann das Effektivitätsmaß natürlich auch dazu verwenden, die Abbildungsgüte zweier Stichproben miteinander zu vergleichen. In diesem Fall zeigt ggf. unterschiedliche Effektivität, daß die ungewichteten Stichproben die Grundgesamtheit unterschiedlich genau abbilden." (Von der Heyde, 1994: 150)

Table 3.7 Efficiencies of the IPF-Solution (in %)

Nation	EB41.0	FORSA
Belgium	92.8	88.4
France	92.6	87.5
Spain	93.9	84.4
East Germany	92.5	69.8
West Germany	94.2	81.9

Table 3.8 Efficiencies of the LS-Solution (in %)

Nation	EB41.0	FORSA
Belgium	92.9	88.7
France	92.7	87.9
Spain	94.0	85.1
East Germany	92.6	73.2
West Germany	94.2	82.4

Table 3.7 and Table 3.8 show higher efficiencies for the Eurobarometer study in all nations with both weighting procedures. We can interpret this result as a sign that the representation of the target population is better achieved in the Eurobarometer study than in the telephone survey. The most similar efficiencies are shown by the two Belgian studies. The largest differences in efficiency between the Eurobarometer study and the telephone survey exist in East Germany. Of course, the weights generated as a result of the LS procedure cause higher efficiencies than the IPF weights for all countries, except for the Eurobarometer study in West Germany were the efficiencies are the same for both solutions.

3.7 Results of the analysis of the impact of the weighting procedures on selected passive variables

In the following section we want to describe the similarities and dissimilarities of selected weighted and unweighted passive variables. For this analysis we used the following indicators (see Chapter 1, for the wording of the questions):

* Satisfaction with life
* Satisfaction with democracy
* Subjective social class identification

These indicators are frequently used in the social sciences.

It is obvious that weighting procedures do not have an impact on the distribution of a passive variable if its relative frequencies are similar for each cell.

For the Eurobarometer study in West Germany Figure 3.1 shows the histograms for the passive variable "Satisfaction with life" with respect to the 32 cells defined by the active variables. The cells are lexicographically ordered with respect to gender, age and size of household. For example, the second picture from the left in the second row is based on the 76

male respondents in West Germany of age 45-59 living in households of size 2. Since most of the histograms are of similar shape we are not surprised that the weighting procedures do not essentially change the distribution of the passive variable. One can also show that the dependency and thus the correlation between our active variables and "Satisfaction with life" is very weak and that therefore the correction is minimal.

It can also be clearly seen in Table 3.9 that weighting has almost no effect on the marginal distributions of the passive variables. To save space we will present only the case of "Satisfaction with life". For the other passive variables mentioned above we obtained similar results.

The profiles are approximately the same before and after weighting but in most cases very different for the Eurobarometer study and the telephone survey.

In order to summarise the similarities and dissimilarities between the weighted and the unweighted data of the two surveys in each nation, we discuss the results of a correspondence analysis shown in Figure 3.2.

The correspondence analysis is a multivariate method for the graphical representation of the rows and columns of a contingency table. The four categories of "Satisfaction with life" serve as rows of the input data matrix. The columns consist of the relative frequencies of the four row categories in the case of the weighted (IPF and LS) and unweighted studies in the different nations.

The circles in the map represent the rows, the squares and triangles the columns. The first letter denotes the nation[15]. P and T, respectively, are abbreviations of Personal Eurobarometer study and Telephone survey, respectively. The squares are used as symbol of the Eurobarometer data, the triangles represent the columns of the telephone survey. The last two letters denote the weighting procedure[16]. The nearer two profiles are the nearer are the corresponding points in the map. Since the quality of the two-dimensional plot is 96.4% we have an excellent representation of the data in the plot. The map shows very clearly that the three points belonging to the same nation and the same survey type are close together. That means the PU and PW as well as the TU and TW of one nation lie side by side, partly one on top of the other. This is an indication that the weighting procedures do not have an impact on the results of the indicator "Satisfaction with life". The reason for it is the similar distribution of this passive variable according to the cells defined by the active variables.

Our map also shows a clear distinction between the Eurobarometer study and the telephone survey. The first axis in the map can be interpreted as the "satisfaction axis". The more to the right the projections of the points onto the first axis are, the more satisfied the respondents on average were. Since the telephone survey points lie always to the right of the corresponding Eurobarometer points, the FORSA respondents in contrast to the Eurobarometer respondents are more satisfied. The difference in Spain is especially striking. In general, this may be an influence of the survey method. To investigate this question further we added two supplementary columns which separate the telephone owners from the respondents without

15 B=Belgium, F=France, S=Spain, O=East Germany, W=West Germany
16 U=unweighted, W=weighted, I=IPF, L=LS

telephone in East Germany. We choose East Germany as an example because of its relative low telephone penetration. The squares labelled by O_PU_TO and O_PU_NTO, respectively, are an indication to the fact that at least in East Germany the answers depend on the telephone ownership. The point O_PU_TO is closer to O_TU. If we proceed in the same way for Spain we obtain a similar effect, although it is not so drastic. The conclusion is that telephone ownership cannot explain the immense difference between the Eurobarometer points and the FORSA points for Spain.

Figure 3.1 Histograms for "Satisfaction with Life" in 32 cells

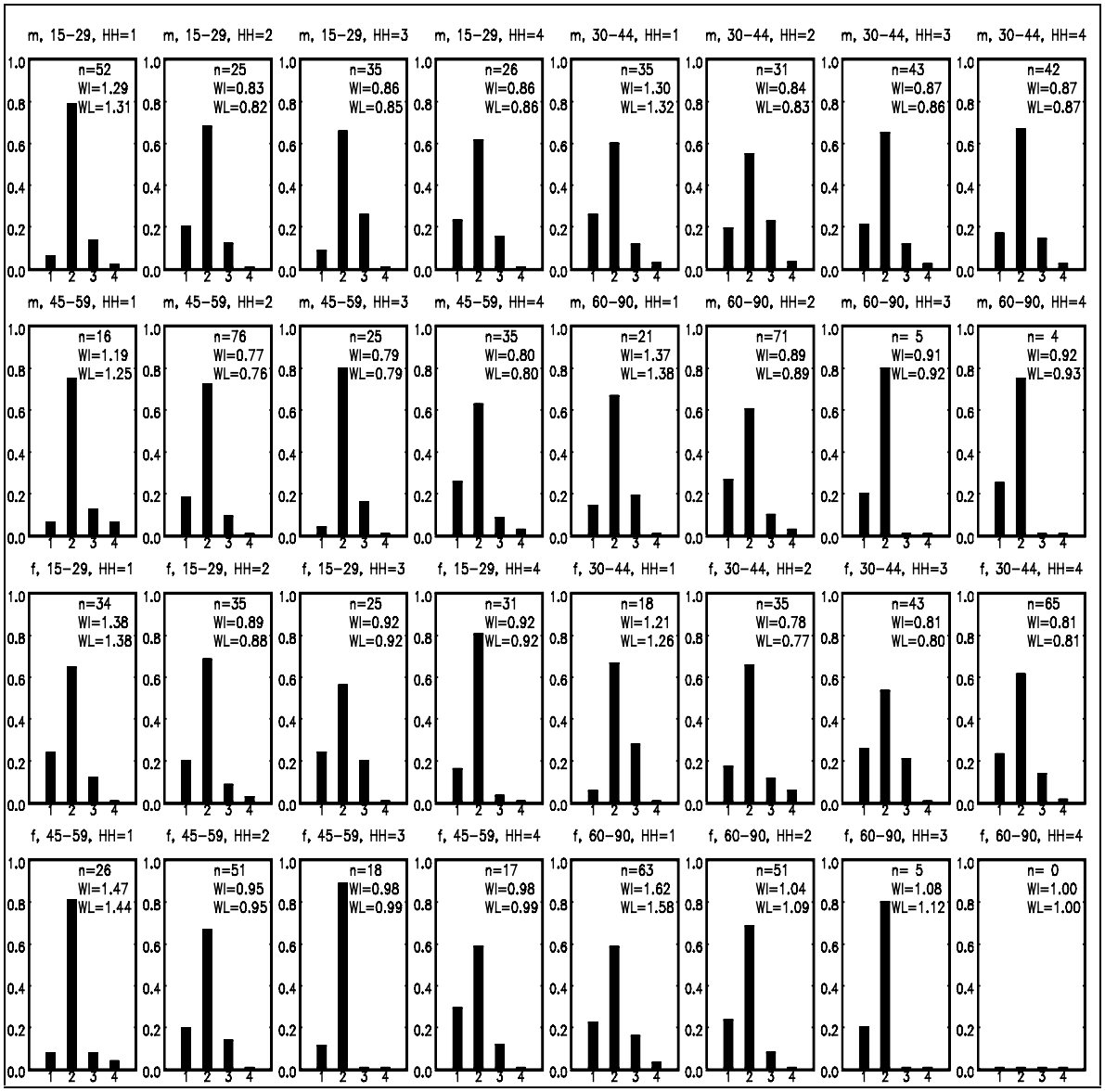


Table 3.9 Satisfaction with life: Marginal distributions for the unweighted and weighted data in five European countries (P=EB41.0, T=FORSA, U=Unweighted, W=Weighted, I=IPF-solution, L=LS-solution)

Satisfaction in:						
Spain	**PU**	**PWI**	**PWL**	**TU**	**TWI**	**TWL**
very satisfied	13.9	13.6	13.6	35.8	36.0	35.9
fairly satisfied	53.8	54.0	54.0	42.8	43.1	43.1
not very satisfied	25.2	25.3	25.3	18.0	17.6	17.7
not at all satisfied	7.1	7.1	7.1	3.4	3.3	3.3
Belgium	**PU**	**PWI**	**PWL**	**TU**	**TWI**	**TWL**
very satisfied	32.5	31.4	31.4	32.7	32.2	32.2
fairly satisfied	56.0	56.3	56.4	60.2	60.0	60.0
not very satisfied	9.0	9.7	9.6	5.3	5.8	5.9
not at all satisfied	2.5	2.6	2.6	1.8	2.0	1.9
France	**PU**	**PWI**	**PWL**	**TU**	**TWI**	**TWL**
very satisfied	13.7	13.3	13.3	21.2	21.9	21.5
fairly satisfied	58.5	58.2	58.3	61.7	61.2	61.5
not very satisfied	21.3	21.5	21.4	11.7	10.7	10.8
not at all satisfied	6.5	7.0	7.0	5.4	6.2	6.2
East Germany	**PU**	**PWI**	**PWL**	**TU**	**TWI**	**TWL**
very satisfied	7.3	7.7	7.7	13.9	13.9	14.0
fairly satisfied	64.5	63.7	63.7	64.1	63.3	64.0
not very satisfied	21.7	21.8	21.8	18.2	18.0	17.2
not at all satisfied	6.5	6.8	6.8	3.8	4.8	4.8
West Germany	**PU**	**PWI**	**PWL**	**TU**	**TWI**	**TWL**
very satisfied	19.1	18.8	18.8	26.7	27.4	27.5
fairly satisfied	66.5	66.6	66.6	61.5	60.1	59.7
not very satisfied	12.9	13.0	13.0	9.4	9.6	9.9
not at all satisfied	1.5	1.6	1.6	2.4	2.9	2.9

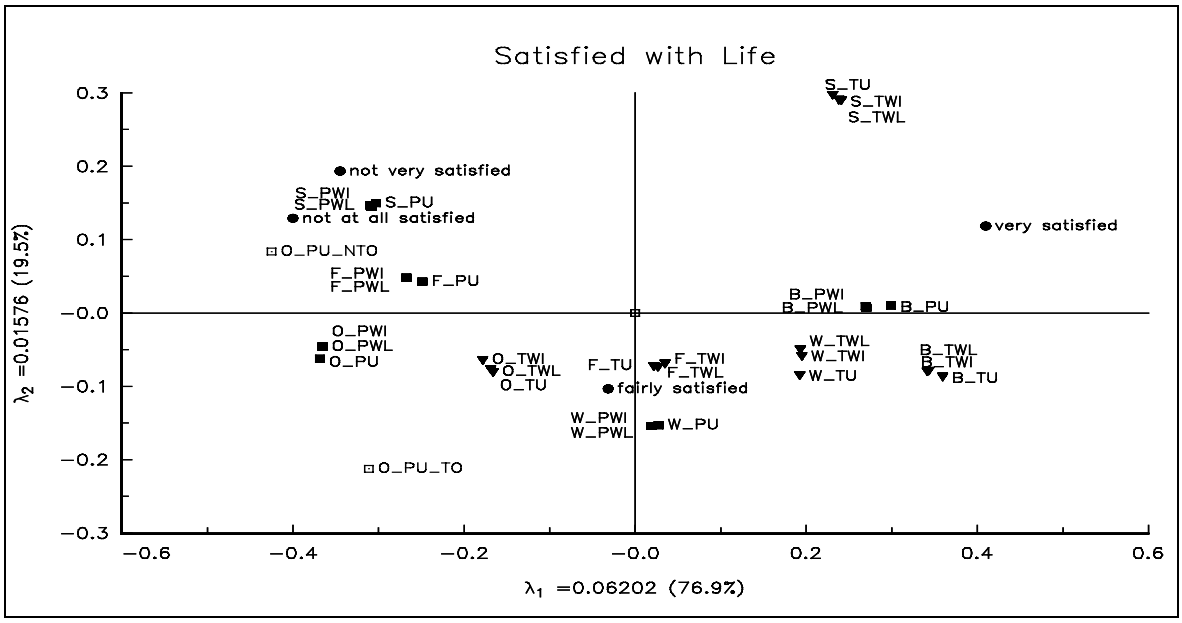
3.8 Conclusion

The results of our analysis lead to the conclusion that both the face to face as well as the telephone sample deviate quite a bit from the different populations with respect to the reference variables. We also found that the weighted and the unweighted passive variables have nearly the same distributions. This means that existing differences between the Eurobarometer study and the telephone survey are not reduced by weighting. This result is valid for all nations, all analysed variables and both weighting routines.

Furthermore, our analysis is in agreement with the findings of Schnell (1993), who looked at the impact of weighting by the active variables gender, age and region on attitude variables of the ALLBUS study. As a summary he stated that it is not possible to derive the unbiasedness

of the passive variables from the unbiasedness of demographic variables in a sample.[17] The reason he gives for this quite consequential conclusion is the only indirect and low impact which the active variables have on the passive variables.

Figure 3.2 Correspondence analytic map of "Satisfaction with Life" for the two modes in the five selected countries



Our analysis also results in the message that cell weighting with the selected demographic variables does not adjust the marginal distributions of the passive variables in both surveys. This means that for our investigation nonresponse bias or sample design bias could not be corrected by the weighting procedures applied.

The crucial problem for the application of weighting procedures seems to be the selection of the active variables. If there exist no relations between the active variables and the survey variables of interest, respectively, weighting does not help. Therefore, a mechanic application of weighting routines without checking these relations does not make much sense.

Furthermore, the results of our analysis show that the Eurobarometer study seems to represent the target population better than the FORSA telephone study does. This could be clearly demonstrated for some demographic variables. There are several indications that this is valid for the selected passive variables, too. However, we cannot generalise this conclusion for all telephone studies since the quality of a survey depends on many factors of which the mode of data collection is only one. Further investigations are necessary to throw light on this problem.

17 Schnell stated "daß aus der Unverzerrtheit 'demographischer Variablen' in den Stichproben nichts über die Unverzerrtheit anderer Variablen folgt" (Schnell, 1993: 29).