Adjustment for differences between face to face and telephone interviews

Saris, Willem E.

CHAPTER 10

ADJUSTMENT FOR DIFFERENCES BETWEEN FACE TO FACE AND TELEPHONE INTERVIEWS

WILLEM E. SARIS

10.1 Introduction

So far the effect of different aspects of the mode of data collection on the results has been studied. Especially, it was shown for the standard questions of the Eurobarometer that the coverage errors moving from face to face to telephone interviewing are relatively small but the mode effects and the effect of the differences in the fieldwork between research organisations can be considerable. In the last part of this book the aim is higher: the aim is to evaluate if it is possible to develop a procedure to make the results of studies done with different modes comparable? In order to do so, this chapter will discuss the statistical basis for this approach and illustrate this for the Eurobarometer experiment done. This approach cannot be directly applied on the Eurobarometer data as collected in this experiment because in the tracking studies of the European Commission a different market research company will do the field work. This requires a similar study as we report here. This chapter starts with a theoretical discussion of the approach. After that the procedures will be illustrated.

10.2 The notation and basic assumptions

In the previous chapters three kinds of problems in comparative survey research have been discussed: coverage errors, nonresponse errors and pure mode errors. The first two kinds of errors are due to a kind of process which will be called ‘selection’. The mode errors are caused by a process which will be called ‘transformation’. These processes can be formulated in very similar ways but nevertheless produce very different results. First, the selection process will be addressed.

10.2.1 Selection processes

One of the simplest selection processes is sampling. In the Eurobarometer and other survey research, people are interested in the distribution of the opinions of people in a population. For example they would like to know: How many people think that their country has benefited or not benefited from the membership in the EU, and to what extent people have no opinion about this topic.

Normally it is assumed that in the population a frequency distribution exists for the opinion one is interested in. That this is an assumption has been elaborated by Zaller (1992). Following Converse (1964), Zaller suggests that people have no fixed opinion about many issues before the interview but create an opinion when they are asked about it. Whether one assumes the existence of an opinion or the creation of an opinion will not change the argument in this chapter.

Whatever assumption is made: The existing or created opinion for a specific question[33] will have a frequency distribution which will be denoted by **f**. Thus **f** contains three numbers for the benefit question, the sum of which gives the total number (N) of people in the population (see table 10.1). This distribution is, of course, not known. One of the purposes of the Eurobarometer studies is to estimate this distribution.

Table 10.1 A potential frequency distribution of the variable "benefit from EU membership"

Opinions	Absolute frequency **f**	Relative frequency **f**/N
Benefited	10.0 million	. 6250
Not benefited	5.0 million	. 3125
DK/No answer	1.0 million	. 0625
Total population	16.0 million	1. 0000

Research can be done in different ways, for practical reasons the population as a whole will hardly ever be used. This means that almost always a sample is drawn from the population at large. In principle the sample of size n should be chosen in such a way that the expected relative frequency distribution of the sample s ($\mathbf{f}_s$/n) is identical to the relative frequency distribution in the population (**f**/N). If the sample size is n, from each class of the population frequency distribution the same proportion of cases should be drawn, namely p_s = n/N. In table 10.2 the example is continued with a sample of size 16.000.

33 The description also covers the ideas of Zaller (1992) who suggests that different considerations exist which lead to a response on the basis of the saliency. If we assume a specific combination of considerations as salient for a specific question, one can represent some aggregated result of these considerations as the opinion of the person at that moment of that question. This is all we need for the formulation in the chapter.

Table 10.2 A potential frequency distribution of the variable "benefit from EU membership" in the sample

Opinions	**Absolute frequency in the population** $\mathbf{f}$	**Expected frequency in the sample** $\mathbf{f_s=p_sf}$	**Relative frequency in both** $\mathbf{f_s/n}$
Benefited	10.0 million	10 thousand	.6250
Not benefited	5.0 million	5 thousand	.3125
DK/No answer	1.0 million	1 thousand	.0625
Total population	16.0 million	16 thousand	1.0000

The selection process does not determine who is chosen but only how many are chosen. If the number in category k is represented by f(k) for the population and by $f_s(k)$ for the sample, we could represent the consequences of this selection process in a relationship between $\mathbf{f}_s$ and $\mathbf{f}$ as follows:

$$f_s(1) = p_s f(1)$$

$$f_s(2) = p_s f(2)$$

$$f_s(3) = p_s f(3)$$

If the probabilities are placed in a diagonal matrix, the outcome of this sampling procedure with equal probabilities can also be presented in matrix notation:

$$\begin{vmatrix} f_s(1) \\ f_s(2) \\ f_s(3) \end{vmatrix} = \begin{vmatrix} p_s & 0 & 0 \\ 0 & p_s & 0 \\ 0 & 0 & p_s \end{vmatrix} \begin{vmatrix} f(1) \\ f(2) \\ f(3) \end{vmatrix}$$

or

$$\mathbf{f}_s = \mathbf{S}_s.\mathbf{f} \qquad (1)$$

where $\mathbf{S}_s$ gives the effects of the selection mechanism. In this case this is a diagonal matrix with equal probabilities on the diagonal. It is essential that the probability of drawing a person from a class is the same for all classes of the variable.

However, such a procedure is very unlikely. There might be coverage errors or nonresponse errors as discussed before. Often such errors are related to the variables of interest. This means that for a specific variable for the members in the different groups, no equal probability exists.

The consequences of such a selection process can be presented in the same way as above, but now with unequal probabilities. For example, the occurrence of coverage errors suggests a selection process with unequal probabilities:

$f_c(1) = p_{c1}f(1)$

$f_c(2) = p_{c2}f(2)$

$f_c(3) = p_{c3}f(3)$

where c stands for coverage and p_{ck} is the probability to end up in $f_c(k)$ coming from f(k). The probabilities are different as a consequence of coverage error. Using matrix notation as before this reads:

$$\begin{vmatrix} f_c(1) \\ f_c(2) \\ f_c(3) \end{vmatrix} = \begin{vmatrix} p_{c1} & 0 & 0 \\ 0 & p_{c2} & 0 \\ 0 & 0 & p_{c3} \end{vmatrix} \begin{vmatrix} f(1) \\ f(2) \\ f(3) \end{vmatrix}$$

or $\mathbf{f}_c = \mathbf{S}_c.\mathbf{f}$ (2)

$\mathbf{S_c}$ is again a diagonal matrix but now with unequal values. As a consequence this matrix $\mathbf{S_c}$ produces a selection ($\mathbf{S}_c$) of the cases in the sample which is biased in some direction. In chapter 5 it was shown that this might occur, for example, if a sample is drawn from telephone owners, and the ownership of the telephone is related to the opinion on the variable of interest.

A similar problem will emerge due to nonresponse. A fieldwork organisation might use a procedure which is such that certain respondents have a higher probability to participate than others. If this selection process is related to the variable of interest bias will occur in the sample. For nonresponse the consequences of this selection process will be denoted by a matrix $\mathbf{S}_n$, and the formulation of the problem is, of course, the same as for coverage errors, i.e.:

$$\begin{vmatrix} f_n(1) \\ f_n(2) \\ f_n(3) \end{vmatrix} = \begin{vmatrix} p_{n1} & 0 & 0 \\ 0 & p_{n2} & 0 \\ 0 & 0 & p_{n3} \end{vmatrix} \begin{vmatrix} f(1) \\ f(2) \\ f(3) \end{vmatrix}$$

or $\mathbf{f}_n = \mathbf{S}_n.\mathbf{f}$ (3)

This selection process is formally analogous to the previous one. Typical for these selection processes is that people keep their score on a variable but they are selected or not in a certain process. So changes in the responses do not occur. In the case of mode effects this is not the case; therefore one can speak of transformation processes.

10.2.2 Transformation processes

The last process to be formulated is the response process. This process has been discussed in chapter 6 and chapter 9. It was suggested that people in, for example, class 1 of the opinion variable not necessarily also say 1 if they are asked for their opinion. This means that they can change their score on the variable. There is possibly a high probability that they say 1 but there is possibly also a nonzero probability that they say 2 or 3. This is not a selection process as discussed above where a person remains in the same class but is selected or not. Rather, here people can move from one class to another. This process can be formulated with a latent class model as before:

$$\begin{vmatrix} f_m\,(1) \\ f_m\,(2) \\ f_m\,(3) \end{vmatrix} = \begin{vmatrix} \pi_{m11} & \pi_{m12} & \pi_{m13} \\ \pi_{m21} & \pi_{m22} & \pi_{m11} \\ \pi_{m31} & \pi_{m32} & \pi_{m33} \end{vmatrix} \begin{vmatrix} f(1) \\ f(2) \\ f(3) \end{vmatrix}$$

or

$$\mathbf{f}_m = \Pi_m \mathbf{f} \qquad (4)$$

In this case **f** represents, as before, the number of people in the classes before the response is given and $\mathbf{f}_m$ the distribution of the answers if mode m is used.

The difference with the selection process is that people in, for example, class 1 have a probability π_{m11} to go to class 1 of the response variable, a probability π_{m21} to go to class 2 and a probability of π_{m31} to go to class 3. In the selection process all probabilities were zero except the probabilities in the diagonal. Therefore, the people will always keep the same score in the selection process. In the response model these probabilities are not zero, and therefore people can move to a different class then they were in before. This is typical for the transformation process.

Above the basic processes were presented which play a role in any survey research.

It should be clear that $\mathbf{f}_c$ and $\mathbf{f}_n$ cannot be observed without asking a question. Thus, also these distributions represent latent classes. Since a response process for the whole population can not be seen as realistic, combinations of the above mentioned processes need to be specified for real-life research.

10.3 Data collection and the assumption of independence

In a face to face interview the following steps are carried out:

1. a specific sample is drawn ($\mathbf{S}_s$)
2. fieldwork is done by organisation *i* leading to a specific nonresponse selection ($\mathbf{S}_{ni}$)
3. data are collected with the face to face mode of data collection (Π_f).

In the above specification a sequence of steps is identified, while in the previous section only single steps were considered. In order to make the formulation simple, one has to assume independence of the different steps.

This means that the following assumptions need to be made:

Assumption 1: The selection within the fieldwork is not different whether the whole population would have been contacted or only a sample.

Assumption 2: The response process can be described by the same response probabilities whether one is concerned with the population at large or a sample or a subsample which is willing to co-operate.

If these assumptions can be made, the resulting frequency distribution of the sequence of steps of the face to face interview ($\mathbf{f}_{ftf}$) can be described as:

$$\mathbf{f}_{ftf} = \Pi_f.\mathbf{S}_{ni}.\mathbf{S}_s.\mathbf{f} \qquad (5)$$

One can read this as follows: The resulting frequency distribution in face to face interviewing ($\mathbf{f}_{ftf}$) will be realised by the sample selection ($\mathbf{S}_s$) from the population distribution (**f**) which is again changed by the selection in the fieldwork($\mathbf{S}_{ni}$) where finally the people give their responses with a certain response probability (Π_f)

In telephone surveys the following steps are taken:

1. a sample is drawn ($\mathbf{S}_s$)
2. from this sample some people drop out because of lack of a telephone ($\mathbf{S}_c$)
3. the fieldwork causes a certain nonresponse selection ($\mathbf{S}_{nj}$) due to organisation j
4. the people answer the questions through the telephone (Π_t).

Using the assumption of independence, the resulting frequency distribution of this telephone ($\mathbf{f}_t$) interview will be:

$$\mathbf{f}_t = \Pi_t.\mathbf{S}_{nj}.\ \mathbf{S}_c\ .\mathbf{S}_s.\mathbf{f} \qquad (6)$$

In this process one additional selection step is necessary due to the fact that not all people have a telephone which might bias the results.

Finally also the panel study of the Eurobarometer experiment should be defined in the same way. This approach started as a face to face study:

1. a specific sample is drawn (S_s)
2. fieldwork is done leading to a specific nonresponse selection ($\mathbf{S}_{ni}$)
3. data are collected with a certain mode of data collection (Π_f)

The data are not used immediately but first some further steps are done in line with the telephone interviewing:

4. from this sample some people drop out because of lack of a telephone ($\mathbf{S}_c$)

5. the people are asked to participate in the panel which causes a certain nonresponse selection ($\mathbf{S}_{pi}$) due to the way organisation i works

6. the people answer the questions through the telephone in the panel (Π_{tp})

If step 3 is ignored for the moment, one can specify this process as follows:

$$\mathbf{f}_{pt} = \Pi_{tp}.\mathbf{S}_{pi}.\ \mathbf{S}_c\ .\mathbf{S}_{ni}.\mathbf{S}_s.\mathbf{f} \tag{7}$$

In this formula the steps mentioned above can be observed: first the selection for the sample, then the selection for the face to face study, then the reduction to telephone owners, and finally the drop out in the panel. The people who are left after all these steps are asked the questions by telephone which leads to the final result denoted as $\mathbf{f}_{pt}$ for the telephone answers of the panel.

For the estimation of all effects one additional assumption is essential:

Assumption 3: The response probabilities in the panel do not differ from the probabilities in a normal telephone or face to face interview.

This assumption is less certain than the previous two assumptions because here one deals with repeated observations and the previous answer can have an effect. However, Van Meurs and Saris (1989) have shown that such effects disappear in most cases after 20 minutes in the same interview so these effects will have most certainly evaporated after one week or more. If this assumption can be made it means that:

$$\Pi_{tp} = \Pi_t \tag{8}$$

and it follows that

$$\mathbf{f}_{pt} = \Pi_t.\mathbf{S}_{pj}.\ \mathbf{S}_c\ .\mathbf{S}_{ni}.\mathbf{S}_s.\mathbf{f} \tag{9}$$

This case equals selection processes as already seen before except for the new selection effect ($\mathbf{S}_{pj}$) due to the use of a panel.

Finally, in chapter 6 the fact was used that for the panel the responses from the face to face interview are available. This results in:

$$\mathbf{f}_{pf} = \Pi_f.\mathbf{S}_{pj}.\ \mathbf{S}_c\ .\mathbf{S}_{ni}.\mathbf{S}_s.\mathbf{f} \tag{10}$$

Now in a more formal way all procedures used in this study are defined. In order to give an idea of the possible consequences of the different processes in survey research, table 10.3 presents the relative frequency distributions for the “benefit” question in France as an example.

Table 10.3 The relative frequency distributions for the “benefit from the EU membership” question in France

Category	Face to face	Telephone
Benefited	39.1	45.0
Not benefited	39.4	30.4
DK/No answer	21.5	24.6
N	1000	500

The table shows clearly that the differences are considerable. It is, however, not clear where these differences come from. Therefore it will be explored in the next section whether the selection and response procedures can be estimated on the basis of the available data and the previous assumptions.

10.4 The estimation of selection and response processes

In the chapter 6 and 10 the latent class model was used to estimate the response process. This was based on the following simplification. In (9) and (10) it can be seen that the selection for the panel leads to the following frequency distribution ($\mathbf{f}_p$):

$$\mathbf{f}_p = \mathbf{S}_{pj}.\ \mathbf{S}_c\ .\mathbf{S}_{ni}.\mathbf{S}_s.\mathbf{f} \tag{11}$$

The resulting distribution is unobserved (latent) because so far no response process is specified. Starting from here there are two modes in which people have responded: face to face and telephone. So one can write instead of (9) and (10):

$$\mathbf{f}_{pt} = \Pi_t\ \mathbf{f}_p \tag{12}$$

and

$$\mathbf{f}_{pf} = \Pi_f\ \mathbf{f}_p \tag{13}$$

Note that it is essential that $\mathbf{f}_p$ is the same for both modes of data collection.

It has been shown in chapters 6 and 10 that the latent class model makes it possible under certain mild assumptions to estimate the response probabilities and the distribution in the latent classes (see chapter 6 for the details).

For example for the "benefit" variable the response probabilities were the same for telephone and face to face research, and the values were as specified in table 10.4.

Table 10.4 The response probabilities for the "benefit from EU membership" question, given the score on the latent variable, estimated from EB41.Panel

Category	Benefited	Not benefited	DK/No answer
Benefited	.8508	.0159	.0147
Not benefited	.0159	.8719	.1197
DK/No answer	.1333	.1121	.8656

These response probabilities were not only the same for telephone and face to face research but also within France, Belgium and Spain. It seems that the errors made in each of the classes were approximately the same in all three countries. However, the distribution over the latent class was different as can be seen in table 10.5.

Table 10.5 The relative frequency distribution in three countries for the "benefit from EU membership" question estimated from EB41.Panel

Country	Benefited	Not benefited	DK/No answer
France	.4589	.4715	.0696
Belgium	.6163	.2511	.1326
Spain	.4676	.4923	.0402

Having shown that the panel data can be used for the estimation of the response process these results will now be used to explore the estimation of the different selection processes. In this context the third assumption plays an important role. Without this assumption no further estimation could be performed.

The coverage errors can easily be estimated so it is reasonable to start with them. $\mathbf{f}_{\mathrm{ftf}}$ includes owners of telephones and nonowners. If the selection effect of telephone ownership ($\mathbf{S}_{\mathrm{c}}$) is applied to this result, one gets:

$$\mathbf{f}_{\mathrm{ftf.c}} = \mathbf{S}_{\mathrm{c}}\mathbf{f}_{\mathrm{ftf}} \qquad (14)$$

Since both frequency distributions can be obtained from the data, $\mathbf{S}_c$ can be obtained as well. It presents the proportions which have to be applied to move from the total sample to the sample of telephone owners. These proportions do not have to be identical for all classes. In case of the “benefit” variable the result is presented in table 10.6.

Table 10.6 The coverage errors in three countries for the variable “benefit from EU membership”in EB41.Panel

Category	France	Belgium	Spain
Benefited	.94	.86	.80
Not benefited	.96	.80	.80
DK/No answer	.91	.77	.76

These data indicate an unequal effects of penetration in the different countries. This is, however, less relevant than the possible biasing effect of the selection by telephone ownership within each country. In this specific case only in Belgium a significant difference between the different categories has been found. In the same way this selection process can be estimated for all variables and countries.

Since the response probabilities are known one can obtain the distribution of the latent opinions of the respondents f_{ni} in face to face research by applying equation (15) on the distribution of the responses in the face to face study.

$$\mathbf{f}_{ni} = \Pi_f^{\ -1}.\mathbf{f}_{ftf} \tag{15}$$

In the same way the distribution of the latent opinion of the telephone respondents f_{nj} can be obtained by equation (16) from the distribution of the responses in the telephone study.

$$\mathbf{f}_{nj} = \Pi_t^{\ -1}.\mathbf{f}_t \tag{16}$$

and for the panel study the distribution of the latent opinion (f_{np}) from the distribution of the observed responses in the panel study by applying equation (17)

$$\mathbf{f}_{np} = \Pi_t^{\ -1}.\mathbf{f}_{pt} \tag{17}$$

The results of these calculations for the first two equations are presented for the “benefit” variable in France in columns 3 and 4 of table 10.7. The resulting frequency distributions represent the estimated frequency distributions of the two studies corrected for mode effects.

Before, it was indicated that these frequency distributions will not be the same because they are effected by different selections processes (coverage errors and nonresponse errors). These selection processes can be written as:

$$\mathbf{f}_{ni} = \mathbf{S}_{ni}.\mathbf{S}_{s}.\mathbf{f} \tag{18}$$

$$\mathbf{f}_{nj} = \mathbf{S}_{nj}.\ \mathbf{S}_{c}\ .\mathbf{S}_{s}.\mathbf{f} \tag{19}$$

$$\mathbf{f}_{np} = \mathbf{S}_{pj}.\ \mathbf{S}_{c}\ .\mathbf{S}_{ni}.\mathbf{S}_{s}.\mathbf{f} \tag{20}$$

From these three equations the effects of nonresponse for the different organisations have to be estimated. Since $\mathbf{S}_c$ is also known, (20) can be used to estimate $\mathbf{S}_{pj}$. This can be done by substitution of the estimated values for $\mathbf{f}_{ni}$ from (15) in (20).

Now, the only remaining task is to estimate from (18) and (19) the selection processes specified by $\mathbf{S}_{ni}$ and $\mathbf{S}_{nj}$ for the normal face to face and telephone studies. It is, however, simple to show that these selection processes cannot be estimated separately.

These equations have the form:

$$f_{nt}(k) = p_{nt}f_s(k) \qquad \text{for } k = 1\text{- } K,\ \ t = 1,2 \tag{21}$$

where k is a category number, t the research organisation and s denotes the sample.

From research 2K numbers $f_{ni}(k)$ are known, but with this information 3K unknowns (2K p_{ni} and K $f_s(k)$ elements) have to be estimated. This is impossible. This also means that one can not get an estimate of f which is the distribution in the population. This means that one has to adjust the aim of the study.

A less attractive result but still very valuable is that one can get the relative size of the different errors. Since this result is also useful it will be presented here although it is not exactly what was wanted.

From (18) and (19) it follows that:

$$\mathbf{f}_{ni} = \mathbf{S}_{ni}.\ (\mathbf{S}_{nj}.\ \mathbf{S}_{c}\ )^{-1}\ \mathbf{f}_{nj} \tag{22}$$

which is in normal algebra:

$$f_{ni}(k) = (p_{ni}/p_{nj}.p_{c.})\ f_{nj}(k) \qquad \text{for each k} \tag{23}$$

which gives :

$$w_{in}(k)= p_{ni}/p_{nj}.p_c = f_{ni}(k)/ f_{nj}(k) \qquad \text{for each k} \qquad (24)$$

Both frequencies can be estimated if the response probabilities are known and p_c is also known. So the ratio p_{ni}/p_{nj} can also be estimated. For the "benefit" example, the results of these calculations are presented in table 10.7.

Table 10.7 The estimates of the nonresponse effects on the "benefit from EU membership" question for the two studies compared (EB41.0 and FORSA)

k	$f_{ltft}(k)$	$f_t(k)$	$f_{ni}(k)$	$f_{nj}(k)$	$w_{ij}(k)$	p_c	p_{ni}/p_{nj}
Benefited	391	450	520	449	.8635	.94	.8117
Not benefited	394	304	317	427	1.3470	.96	1.2931
DK/No answer	216	246	163	125	.7669	.91	.6969

In this case the organisation which did the face to face interviews has reached relatively many people with a negative opinion, and the other company which organised the telephone survey obtained co-operation of relatively many respondents with a positive opinion or no opinion at all (DK/No answer). More cannot be said about these differences on the basis of the data. With respect to the last point it should be made clear that in the comparison between the (non) response of the different companies the mode effects do not play a role any more because a correction was already made for this factor.

Although these results are interesting in itself they are not what was desired. So far there is no possibility to estimate the size of the errors and, therefore, no correction can be made for them. Therefore, in the next section an alternative will be formulated.

10.5 Prediction of the face to face results from telephone data

Since the estimation of all errors is not possible, the aim should be to obtain at least a procedure to predict the face to face results from the telephone data or vice versa. If this is possible one can use one mode of data collection to report about the other mode. In this way one can avoid differences in the reporting.

In case a panel study is done using the two modes it seems obvious that one can use the "turnover table" giving the relationships between the responses in the different modes for this correction. As an example table 10.8 presents such a table which has been produced with the latent class proportion of .9 and .1 and unequal response probabilities for the face to face (A) and telephone mode (B).

This table shows the distributions of the two variables in the marginals while the combinations of the values on A and B are in the cells. Within brackets column percentages are presented which could be used to compute the distribution of the variable A if the distribution of variable B is obtained. With the row percentages of the table one could create the distribution of the variables B from the distribution of the variable A. This result seems to suggest that this table can be used to estimate the distribution of A from B or the distribution of B from A.

Table 10.8 The relationship between responses in A and B if $\pi_1^x = .9$ and $\pi_2^x = .1$

	Variable B 1		2		Total
Variable A					
1	.652	(.767)	.078	(.52)	.730
2	.198	(.232)	.072	(.48)	.270
Total	.85	(1.000)	.150	1.000	1.000

There are two objections against this idea. The first concerns possible changes in the latent classes. It is indeed true for the given data that the turnover table can be used, but if one would like to use the same turnover table which has been obtained at some point in time at a different point in time this procedure is quite doubtful unless the distribution of the latent variable x has not changed. If the distribution of this variable has changed, one should use a different turnover table even if the response probabilities remained exactly the same.

This point will be illustrated by an example. Imagine that the only difference with the previous example is that the people have changed their opinions. Now $\pi_1^x = .7$ and $\pi_2^x = .3$ and not 9 and .1 as before, while the response probabilities remain the same as before. Then quite a different turnover table is obtained (Hagenaars, 1994) which is shown below.

Table 10.9 The relationship between responses in mode A and B if A and B if π_1^x = .7 and π_2^x = .3

		Variable B				
		1		2		Total
Variable A	1	.515	(.69)	.074	(.31)	.590
	2	.235	(.31)	.166	(.69)	.410
	Total	.750	(1.00)	.240	1.00	1.00

This table shows that the probabilities which should be used in this table to calculate the distribution of A from the distribution of B are very different from the previous table even though the only difference is the distribution of the latent variable. This means that this table cannot be used for these calculations because this transformation is needed at different points in time, and at each occasion one can expect changes in the opinion. So equality of opinion cannot be assumed.

The second objection is that using this table corrects only pure mode effects, and it was shown in chapter 5 that the nonresponse effects are often at least as large. But non response effects are ignored in this approach. Therefore, one has to use a more complex approach.

Although the simple estimation procedure using the turnover table is not possible, the turnover table obtainable by panel data is nevertheless useful because it can be used to estimate the response probabilities. If these probabilities remain stable, which is much more likely than the stability of the distribution of the opinion, an estimate of the distribution of the latent variable from the distribution of the observed variables is possible. Combining these results with the results of the previous section, a correction procedure can be formulated.

This can be done starting with equation (6). From (6) follows:

$$\mathbf{S}_s.\mathbf{f} = (\Pi_t.\mathbf{S}_{nj}.\ \mathbf{S}_c)^{-1}.\ \mathbf{f}_t \tag{26}$$

and substitution of this result in (5) gives

$$\mathbf{f}_{ftf} = \Pi_{\mathbf{ftf}}\ .\mathbf{S}_{ni}.\ (\Pi_{\mathbf{t}}\ .\mathbf{S}_{nj}.\ \mathbf{S}_c\ )^{-1}.\mathbf{f}_t \tag{27}$$

which is the same as

$$\mathbf{f}_{ftf} = \Pi_{\mathbf{ftf}}\ .\mathbf{S}_{ni}.\ (\mathbf{S}_{nj}.\ \mathbf{S}_c\ )^{-1}\ \Pi_{\mathbf{t}}^{-1}.\mathbf{f}_t \tag{28}$$

and simplifies to

$$\mathbf{f}_{\mathrm{ftf}} = \Pi_{\mathbf{ftf}} \,.\, \mathbf{W}_{\mathbf{ij}} \,.\, \Pi_{\mathbf{t}}^{-1}.\mathbf{f}_{\mathbf{t}} \tag{29}$$

where **W**ij is a diagonal matrix with as elements on the diagonal the values w_{ij} which represent the relative effects of the different organisations on nonresponse and response in the different categories of the variable (including the coverage error). In the last section it was shown that these coefficients can be estimated (23).

Since the response probabilities and the weights are known, this equation can be used for estimating the face to face results from the telephone results even if the studies are done by different companies. For the "benefit" variable the calculations are illustrated in table 10.10.

Table 10.10 The estimation for France of the face to face results from the telephone data for the variable "benefit from EU membership"

Categories	$\mathbf{f}_\mathbf{t}$	$\mathbf{f}_\mathbf{nj}$	$\mathbf{w}_\mathbf{ij}$	$\mathbf{f}_\mathbf{nj}$	$\mathbf{f}_\mathbf{ftf}$
Benefited	450	520	0.8633	449	391
Not benefited	304	317	1.3470	427	394
DK/No answer	246	163	0.7669	125	216

According to equation (29), first $\mathbf{f}_\mathbf{t}$ is corrected for mode effects using $\Pi_\mathbf{t}^{-1}$ to obtain the distributions of the latent variable for the telephone survey. Next the nonresponse and coverage errors are corrected using $(\mathbf{W}_\mathbf{ij})$ so that the latent variable for the face to face survey is obtained. Finally, the results have been made comparable by applying the mode error $\Pi_\mathbf{t}$ of the face to face study on this latent variable in order to get an estimate of the frequency distribution of the face to face study ($\mathbf{f}_{\mathrm{ftf}}$).

In this case, it should not come as a surprise that the results are exactly correct because all estimates are based on the same data and determined by these data. The real test can only be done with new data where the response probabilities of this study are used and the nonresponse weights are obtained from a comparison of two companies who are doing the standard Eurobarometer study and the tracking study. However, this theoretical analysis shows that a prediction from the telephone data to the face to face data is possible.

10.6 Conclusion

In this chapter, first the consequences of research designs for response distributions were formally defined. In doing so it was shown that differences in results can come from selection processes like sampling, coverage errors, nonresponse errors and from transformation processes like response processes.

Next, an effort was made to estimate the potentially biasing factors which turned out not to be completely possible in this experimental design. The response probabilities could be estimated, as could the coverage errors, but the nonresponse selection process for the two different modes could not be estimated separately. Only a ratio of the effects of the two procedures could be assessed.

Furthermore it was shown that the possibility to estimate the response probabilities, the coverage errors and the ratios of the nonresponse errors is enough to estimate the face to face frequency distribution from the distribution in the telephone survey.

It should be remarked, however, that for this the weights (ratios) have to be estimated for all variables separately because they can be different for all variables, as was the case for the response probabilities.

Furthermore, it is also required that the procedures of the research companies doing the research are not changing. If a change happens, the correction factors will probably also change, especially those factors which correct for nonresponse error.