2004/3 March 2004 Documents

Statistics Norway
Department of Economic Statistics

Jan Henrik Wang

Non-response in the Norwegian Business Tendency Survey

Preface

This paper describes methods for adjusting for non-response set in the empirical framework of the Norwegian Business Tendency Survey for manufacturing, mining and quarrying. The paper has previously been published in Norwegian in Statistics Norway's publication Notater, no. 2003/81. In the analysis we use some well-known techniques for adjusting for non-response in sample surveys. Different models for weighting for non-response and methods of imputation are investigated and compared.

Contents

1. Introduction

This paper provides an empirical analysis of different methods for adjusting for non-response in the Norwegian Business Tendency Survey for manufacturing, quarrying and mining (BTS).

The survey maps out manufacturing leaders' judgement of the business situation and the outlook for a fixed set of indicators such as level of production, capacity utilisation, employment and judgement of the general outlook. The survey was developed in 1973 and put into operation on a regular basis from the first quarter of 1974.

Chapter 2 describes the survey's analytical framework and defines the variable of interest and the period of analysis. Further, chapter 3 we perform an analysis of non-response in the Business Tendency Survey by the use of different techniques for adjusting for non-response. A summary of the results will be presented in chapter 4.

2. About the survey

2.1 Unit, scope and sample

The reporting unit is defined as a *kind of activity unit* (KAU). The KAU is derived by combining all establishments in an enterprise carrying out the same industrial activity, regardless of the location of the activity. The industrial activity is defined as 3-digit industry group (SIC94) – in the following referred to as the industry. In the data collection process the response unit is defined as the largest establishment in the KAU, i.e. highest number of employees, but other response units may also be used, such as the enterprise head office. These adjustments are usually adapted in accordance with the wishes of the enterprise, but may also occur for other reasons. In analysis and in the calculation of the statistics the unit is defined as the KAU.

The population covers all KAUs within Mining and quarrying (10, 13-14) and Manufacturing (15-37), see the Standard of Industrial Classification 1994 (SIC94). The Business and Enterprise Register defines the population. The sampling population does not normally include units with 10 employees or less. The sample frame is defined by a status file in the second quarter of every year, and the sample is updated once a year.

The new sampling plan – introduced in the first quarter of 1996 – was designed with the purpose of achieving a holistic overview of the business tendency situation and outlook for each industry[1]. The KAU's employment is used as a measure of size in the stratification process in connection with the sampling of units, where each industry population is divided into four strata.

Stratum 1	Units with 300 employees and over
Stratum 2	Units with 200 - 299 employees
Stratum 3	Units with 100 - 199 employees
Stratum 4	Units with less than 100 employees

Units with more than 300 employees are included as a panel (stratum 1). Units are drawn with a probability proportional to its size in the remaining strata (proportional allocation). This process of drawing the sample is carried out for each stratum in each industry.

[1] The sampling plan was adjusted in the second quarter of 1997. Some adjustments were made to the stratum classification and a part of the original sample was removed and replaced by new units. The sample was also supplemented and the size of the gross sample became approx. 700 units.

In the analytical part of this paper we have simplified somewhat by assuming that the probability of being drawn is the same within each stratum, and that the probability depends on the coverage of employees drawn in each stratum. This is done to simulate the fact that there is an over-representation of the larger units in each stratum.

The gross sample covers approximately 54 per cent of total employment in the population and some 62 per cent of total turnover. The coverage, however, varies between industries. On 2-digit industry group level the coverage varies from 30 to 90 per cent. However, in some industries the coverage could be higher and lower than this.

2.2 Calculation of statistics and weighting of answers

2.2.1 On stratum level

Results on stratum level are calculated by assigning each active unit's answer a weight equal to its employment. More precisely, the calculation of the share of answers in per cent, $SY_{n,i,j,B}$, for question *n*, response alternative *i*, in stratum *j* and industry *B* may be expressed by the following three steps:

The number of employees coded to response alternative *i* is:

$$(1) \qquad Y_{n,i,j,B} = \Sigma_b (\alpha_{b,j} * \beta_{b,i} * S_{b,j,B})$$

where

- $\alpha_{b,j}$ states if a unit is included in the sample in stratum *j*, and if it is active, i.e. has answered the questionnaire in the relevant quarter. $\alpha_{b,j}$ may adopt the values 0 / 1. An active unit is given a value equal to 1 - otherwise 0. A unit that is not represented in the sample is given the value 0 in the calculation of the stratum level.
- $\beta_{b,i}$ may take the values 0/1, depending on which response alternative the individual respondent in the stratum has chosen for the relevant question. For instance, if a respondent has chosen «increase», it is given a factor equal to 1 when the share of answers for this response alternative is calculated. Otherwise the value 0 is used.
- $S_{b,j,B}$ expresses the employment for the individual KAU, *b*, in the stratum population, *j*, in industry *B*.

The total number of employees for all active KAUs in stratum *j* is:

$$(2) \qquad SS_{n,i,j,B} = \Sigma_i \Sigma_b (\alpha_{b,j} * \beta_{b,i} * S_{b,j,B})$$

The share of answers in per cent for alternative *i,* in stratum *j* is then expressed by:

$$(3) \qquad SY_{n,i,j,B} = Y_{n,i,j,B} * 100 / SS_{n,i,j,B}$$

(1) - (3) show that the basis for the calculation of the proportion of answers for a valid response alternative for question *n* is all KAUs which are coded to the same industry in the population. A unit which is not part of the sample or which is not active (non-response) in a quarter is removed from the calculation by the use of the α-factor. The β-factor is used to group the response alternatives that the active units have chosen, and is given a weight equal to the KAU's level of employment.

It follows from (1) - (3) that the sum of share of answers in per cent for a question is equal to 100, i.e.:

$$(4) \qquad \Sigma_i SY_{n,i,j,B} = 100$$

2.2.2 On aggregated levels

Calculations of the proportion of answers on industry level are based on the proportion of answers on the stratum level. In the transition from stratum to industry, the stratum results are, however, weighted with the population employment to correct for relative differences between the strata in a particular industry. More precisely, the calculation of the share of answers in per cent, $SY_{n,i,B}$, for question *n*, response alternative *i,* in industry *B* may be expressed by the following equations:

$$(5) \qquad SY_{n,i,B} = (\Sigma_j \, Y_{n,i,j,B} * a_{j,B}) * 100 / SS_B$$

where SS_B is the sum of employment for all units in the individual stratum population in industry *B*

and

$$(6) \qquad a_{j,B} = 1 / (SS_{n,i,j,B} / SS_{j,B})$$

Equation (6) expresses the inverse of the sum of the probability to be drawn for active units in stratum *j,* industry *B*.

The share of answers in per cent for alternative *i* on the industry level emerge by adding up the product of the number of employees allocated to each response alternative in stratum *j* with the inverse sum of the probability to be drawn for active units in the stratum.

The same principles apply to further aggregation.

As this explanation of the calculation of the proportion of answers and the weighting of replies in the BTS shows, the share of answers for the net sample is calculated in each stratum before the population share is calculated by weighting the inverse *sum* of the probability to be drawn for units in the net sample in stratum *j,* industry *B*. To be able to conduct the analysis of non-response, we must use the inverse probability to be drawn as design weight for each unit, and then aggregate. The calculation procedures used in this paper will therefore deviate somewhat from the procedures used in the quarterly production of the statistics.

2.3 Variable of interest

The questionnaire for the Norwegian Business Tendency Survey contains 28 questions on different aspects of the observation unit's market situation. To simplify the analysis we have only used one of the questions in the survey: *General judgement of the outlook for the establishment in the next quarter*[2]. This question has three response alternatives:

- *Better*
- *Unchanged*
- *Worse*

We have further defined the response alternative as 1 if the unit has answered *'better'* and 0 if one of the other alternatives has been chosen. Because the answers are weighted with the KAU's employment, the variable of interest used in the analysis of non-response is defined as the response alternative multiplied with the unit's level of employment.

In the calculation procedures used in the quarterly production of the statistics a share of answers is calculated for each of the three response alternatives. In addition, a share of the net sample that has not

[2] This is question 18 in the questionnaire and the complete text for the question is: How do you judge - generally for the establishment's business situation in this industry - the outlook for the forthcoming quarter compared with the situation in the present quarter?

answered the question (Item non-response[3]) is computed. From these results balances and diffusion indices are computed for the various questions and industries.

Balance = Positive - negative

Diffusion index = Positive + (0,5*neutral)

2.4 Period of analysis

We use data from the survey carried out in the second quarter of 2003. The table below shows the number of units in the population and sample in the different employment strata. There were a total of 24 438 units in the population and a gross sample of 701 KAUs.

Table 1 Population and sample

Employment stratum	Population	Gross sample
300 or more	159	146[4]
299 - 200	75	38
199 - 100	275	143
99 - 1	23929	374
Sum	24438	701

[3] See Chapter 3 for more information on item non-response.

[4] As the table shows, not all units in the stratum '300 or more employees' are included in the gross sample even though the probability to be drawn is 1 (see chapter 2.1). This is because some units have reported that they do not wish to participate in the survey and consequently have been removed from the sample.

3. Non-response in the Business Tendency Survey

Participation in the Business Tendency Survey is voluntary and we therefore experience a somewhat higher share of non-response than in compulsory surveys. If we look at other surveys aimed at the same population (manufacturing, quarrying and mining), for instance the Norwegian quarterly investment statistics or the statistics on new orders, which both are compulsory, the response rate is close to 98 per cent.

Unit non-response in the Business Tendency Survey – i.e. units in the sample that have not returned the questionnaire – is quite stable at around 15 per cent. This corresponds to an average response rate of 85 per cent in recent quarters.

Item non-response – i.e. missing values for some, but not all of the questions on a questionnaire – varies between the different questions. A summary of item non-response in the second quarter of 2003 shows that it varies between 9.7 and 0.1 per cent. The reason for the huge variation between different questions is that some questions are not relevant for all industries and are not answered. However, the questions of focus in the press release have a low level of item non-response.

There may be different sources and causes for non-response in the Business Tendency Survey. As mentioned earlier the survey is voluntary and some respondents choose not to participate in the survey. The sample is based on a panel where units that have gone bankrupt or have shut down are replaced with new units annually. In addition, units that have not replied in the previous two quarters are removed and replaced by new units drawn from the population using proportional allocation. The survey is based on postal questionnaires. The following causes may be identified as reasons for non-response:

Unit non-response:

- Does not want to participate. Most respondents in the sample based on the population of units in manufacturing, quarrying and mining also participate in compulsory surveys. Some units therefore choose not to reply because the survey is voluntary and the response burden is considered too high.
- Questionnaire does not reach the contact. In some cases the contact person has either left the company or is not present, and therefore the questionnaire does not reach the right person.
- The questionnaire has not been printed for all units.
- The questionnaire is not registered in the data collection process.
- Error in the calculation process. Registered questionnaires are not included in the calculation of the aggregates.

Experience shows that reluctance to participate is the biggest reason for non-response. To control whether the questionnaire is sent to the right place and person, cross comparisons are carried out between units in the sample of the Business Tendency Survey and samples of other surveys with the same population. In most cases of non-response we receive the questionnaire for the compulsory surveys, but not the one for the Business Tendency Survey, even if the respondent is the same.

Item non-response:

- Irrelevant question. The same questionnaire is sent to all respondents, irrespective of which industry they are in. This may lead to problems in answering all the questions in the questionnaire for respondents in some industries. Introducing the response alternative 'Not relevant' has reduced this problem. But because it is believed to be tempting to use this alternative too often, it is not included for all questions.
- Wrong contact. The questions in the Business Tendency Survey require that the respondent has thorough knowledge of a number of economic variables connected to the establishment's activity. This is not always the case, and thus we may experience item non-response.

- The question is not understood. The respondents may not understand all the questions and so do not answer them.
- Error in the registration process. In most cases the questionnaires are optically read and transferred to an electronic medium. In such cases errors are not common. However, questionnaires that cannot be verified optically (fax, copy) are manually registered. In this process it may occur that some questions are not registered.

For the question used in this analysis, *General judgement of the outlook for the establishment in the next quarter*, the choice of answers and non-response in the second quarter of 2003 are distributed as shown in table 2.

Table 2 Choice of response alternative and non-response in the various employment strata

Employment stratum	Better	Unchanged	Worse	Net sample	Item non-response	Unit non-response	Gross sample
300 and over	27	79	19	125	1	20	146
299 - 200	7	18	9	34	1	3	38
199 - 100	25	69	32	126	1	16	143
99 - 0	84	166	74	324	1	49	374
Sum	143	332	134	609	4	88	701

Table 2 shows that item non-response is evenly distributed with one unit in each strata, and amounts to a non-response rate of 0.6 per cent in relation to the gross sample. In the further analysis we consider the item non-response together with the unit non-response. This means that the total non-response is 92 units. This level of non-response produces a response rate of 86.9 per cent. A closer look at the non-response in each employment stratum reveals the following response rates:

Table 3 Response rates

Employment stratum	Response rate
300 and over	85.6
299 - 200	89.5
199 - 100	88.1
99 - 0	86.6
Total	86.9

3.1 Adjusting for non-response

There is no large variation in the employment stratum non-response rate (Table 3). In the current estimation of the Norwegian Business Tendency Survey, it is assumed that the non-response is missing-completely-at-random (MCAR). Non-response is imputed implicitly by treating the net sample estimates as the gross sample estimates. In the following analysis we will take a closer look at this assumption to see if it holds, or if it is better to use a more complex modelling of the non-response.

The first part of the analysis uses different non-response models (weighting) to adjust for non-response. In the second part we test different methods of imputation of the non-response units.

The following notation will be used:

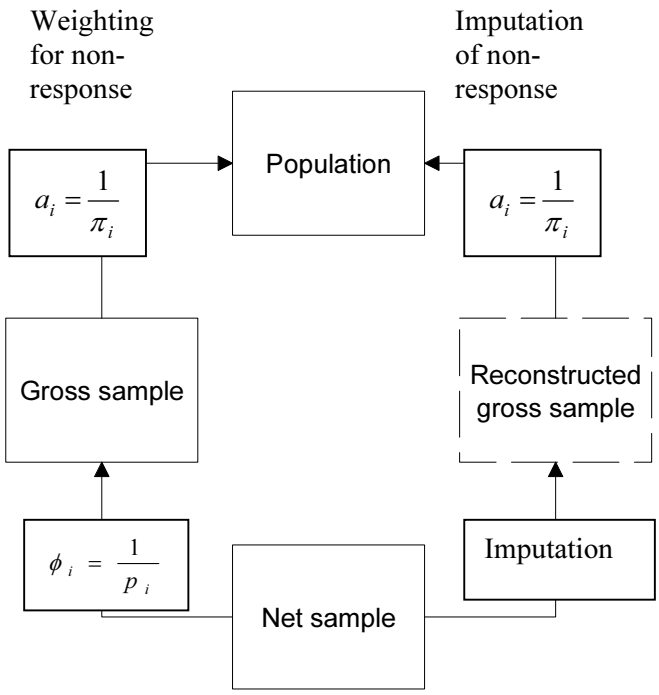
- U = {1,...,N} => Population & i = unit index

- s = (gross-)sample & s_r = net sample (reply sample) & s_m = unit non-response
- r_i is the response variable => r_i=1 if $i \in s_r$ & r_i=0 if $i \in s_m$

- π_i is the probability to be drawn & $a_i = 1/\pi_i$ => Design weight
- p_i is the response probability & $\phi_i = 1/p_i$ => Non-response weight
- $w_i = a_i\phi_i = (\pi_i p_i)^{-1}$ => Non-response-adjusted weight for $i \in s_r$
- y_i is the variable of interest & $Y = \sum_{i \in U} y_i$ => total of y_i in the population

The figure below illustrates the difference between weighting and imputation

Fig 1 Weighting and imputation



As the figure shows, the difference between weighting and imputation is that with weighting we inflate (analogous to the step from gross sample to population level) the sample from net to gross before we calculate the population level, while in imputation missing answers are replaced with estimated values before calculating the population level. When weighting, the product of design weight and non-response weight define the non-response-adjusted weight:

$$w_i = a_i\phi_i = (\pi_i p_i)^{-1}$$

In the analysis that follows, we look at the proportion who respond that the general outlook is *better*. We simplify by only calculating aggregated results for manufacturing, quarrying and mining and not for each industry.

To undertake a structured implementation of the different models and methods of imputation, data at unit level have been adapted in SAS software, and all calculations are carried out in SAS.
The variable of interest in the analysis is defined as

(1) $y_i = \beta_i * S_i$

Where $\beta_i = \begin{cases} 1 & \text{If unit } i \text{ has chosen } \textit{'better'} \\ 0 & \text{If unit } i \text{ has chosen a different response alternative} \end{cases}$

S_i is the unit's number of employees

We want to estimate the proportion of employees that consider the general outlook to be better for the forthcoming quarter, $\bar{Y}$, described by equation (2)

(2) $\bar{Y} = (\sum_{i \in U} y_i) / \sum_{i \in U} S_i$

From the population file we find that the total level of employment is $S = \sum_{i \in U} S_i = 292940$

3.1.1 Weighting for unit non-response

3.1.1.1 Direct weighting

In this section we assume that the non-response is MCAR and use the method of direct weighting. We consider the non-response as an additional phase in drawing a probability based sample. The inverse response probability is used as a non-response weight. Thus, the non-response-adjusted weight is the product of the design and non-response weights.

An estimate for $\bar{Y}$, as the proportion of the employment-weighted answers for the units who have answered *'better'* on the question of the general outlook, may then be expressed as

(3) $\hat{\bar{Y}} = (\sum_{i \in s_r} w_i y_i) / (\sum_{i \in U} S_i)$

To find w_i we have to estimate the response probabilities, p_i, and the non-response weights, ϕ_i, in such a way that we can estimate the non-response-adjusted weight defined by (4):

(4) $w_i = a_i \phi_i = (\pi_i p_i)^{-1}$ where $\phi_i = (p_i)^{-1} = \left(\frac{n}{n+m} \right)^{-1}$

and n is the number of units in s_r, and m the number of units in s_m.

By using direct weighting with MCAR non-response ϕ_i will be constant, i.e. that the response probability is the same regardless of which unit we look at. With these assumptions we get the following estimate

(3) $\hat{\bar{Y}} = (\sum_{i \in s_r} w_i y_i) / (\sum_{i \in U} S_i) = \frac{68372.7}{292940} = 0.233$

Which means that within manufacturing, quarrying and mining 23.3 per cent consider the general outlook for the forthcoming quarter to be better.

3.1.1.2 Estimation with a non-informative RHG-model

We will now try to estimate the share of answers using a non-informative RHG[5]-model. With this type of model we try to divide the sample into groups that are believed to have the same mechanisms for generating non-response. This model is designed to adjust for variation that arises because the non-response is considerably higher within certain groups of the sample. These groups may be defined as units within the same employment strata or within the same industry or other constellations where you expect that the non-response may be correlated with the composition of the groups. The aim of dividing into these response homogeneity groups is to generate a response probability p_i that is as equal as possible within each group, and at the same time as unequal as possible between the different groups. In general the model may be expressed in the following way:

- We assume that the sample is divided into *G* RHGs, defined by s_g for $g = 1,...,G$. Let s_{rg} include response units in s_g, and let s_{mg} include non-response units in s_g in such a way that $s_g = s_{rg} \bigcup s_{mg}$
- We let n_g be the number of units in s_{rg}, and m_g the number of units in s_{mg}. We can then estimate the response probability, p_i, for $i \in s_g$ as described by equation (5)

(5) $$p_i = n_g /(n_g + m_g)$$

By using (5) in (4) we can calculate the non-response-adjusted weight to each unit dependent on which RHG the unit is classified under. Further aggregation is carried out as described in (3).

Within this framework we have chosen to explore two possible classifications of the response homogeneity groups. In *a)* we have grouped units within the same employment strata, and in *b)* we have divided the sample into groups depending on which industry the units are classified under.

The assumption under *a)* implies that there is a higher probability of non-response among the smaller units than among the larger. However, from table 3 above, there is no indication of a considerable difference in the rate of non-response among the smaller and larger units, where size is defined by number of employees.

The assumption under *b)* implies that the rate of non-response may be higher within some industries, and because of this there is a systematic variation in the rate of non-response.

a) Grouping by employment strata

By setting an RHG-index equal to the variable for employment strata in the program used for estimation of the direct weighted estimate with MCAR non-response, we obtain the estimate based on the non-informative RHG-model. The model will contain four response homogeneity groups that correspond to the employment strata shown in table 4 (*g=1,2,3,4)*.

Table 4 RHG = Employment strata

RHG	Employment strata
1	300 and over
2	299 - 200
3	199 - 100
4	99 - 1

[5] RHG = Response homogeneity group

This model produces the following estimate

$$(3) \qquad \hat{\bar{Y}} = (\sum_{i \in s_r} w_i y_i)/(\sum_{i \in U} S_i) = \frac{68646.2}{292940} = 0.234$$

From the estimate we see that it is approximately equal to the assumption of MCAR non-response. This is no surprise when the response rate in the different employment strata was approximately the same (see table 3).

b) Grouping by industry
In this case we chose to group the response homogeneity groups according to which industry the unit is classified under. To avoid too many groups the units are grouped according to publication level. Table 5 shows the coherence between NACE 2-digit level and stratum. The table also includes the response rate in each stratum.

Table 5 RHG = Industry

RHG	Industry[6]	Response rate
1	10, 13 -14	91,7
2	15 - 16	85,5
3	17 - 19	76,7
4	20	89,3
5	21	95,5
6	22	89,2
7	23 - 24	87,1
8	25	81,0
9	26	81,8
10	27	95,5
11	28	93,1
12	29	77,6
13	30 - 33	88,5
14	34 - 35	83,6
15	36 - 37	92,5

Table 5 shows that the response rate varies between the different RHGs and especially that *g=3* (NACE 17 -19; Textiles, wearing apparel and leather) and *g=12* (NACE 29; Machinery and equipment) have lower response rates than the other RHGs.

With this model we get the same estimate as under the model with RHG = Employment strata:

$$(3) \qquad \hat{\bar{Y}} = (\sum_{i \in s_r} w_i y_i)/(\sum_{i \in U} S_i) = \frac{68643.0}{292940} = 0.234$$

The program used in the calculation is the same as under *a)* except that we have changed the RHG-index from *g=x* with *g=x2* (*g* defining the RHG-index, x is employment strata and x2 is the indexation of the industry groups).

From these calculations there does not appear to be a clear correlation between the RHGs we have defined and the non-response, at least not in such a way that it affects the estimate.

[6] The numbers in the column correspond to 2-digit NACE (SIC94)

As shown in table 5, some industries have a lower response rate than others, but the proportion of employment varies significantly between the different industry groups. If we look at the group Textiles, wearing apparel and leather; *g=3*, this group will have a higher numerical value on the non-response weight than the other groups. However, this will be of little significance on the total employment-weighted estimate because this group has a very small proportion of the overall employment in Norwegian manufacturing industry.

3.1.1.3 Estimation with a simple informative RHG-model

This model assumes that the non-response is correlated with the variable of interest. I.e. that it is assumed that the non-response is higher or lower among units choosing one response alternative above another.

- we define RHGs s_g for $g = 1,...,G$ that, among other things, depend on the variable of interest
- further, *auxiliary groups* s_h for $h=1,...,H$ are generated based on variables that are known in the entire sample
- suppose further that the response probability to unit *i* is independent of $i \in s_h$ given that $i \in s_g$

On these assumptions we suppose that the non-response is homogeneous among respondents who answer *better, unchanged* or *worse* and that the response alternatives define s_g for *g=1,..,3.*

Because the variable of interest is unknown for the non-response units, s_{mg}, we have to estimate which group they belong to. To do that, auxiliary groups based on variables known in the entire sample are used. We therefore assume that we have the auxiliary groups s_h for *h=1,..,4*, defined by the four employment strata which are known for the entire sample, $s = s_r + s_m$. Finally, we assume that the non-response is independent of the employment strata given the response alternative.

To estimate the response probabilities we also have to estimate which group the non-response units belong to.

- We let s_{rgh} denote the part of the sample $s_{rg} \cap s_{rh}$, i.e. response units which belong to both s_g and s_h. Further, we let s_{mgh} denote the part of the sample $s_{mg} \cap s_{mh}$, i.e. non-response units which belong to both s_g and s_h.
- We denote the size of s_{rgh}, which is known in the entire sample, with n_{gh}. Further, we let m_{gh} be the size of s_{mgh}, which is unknown except from $m_h = \sum_{g=1}^{G} m_{gh}$ since s_h is known.
- Given an estimate for m_{gh}, denoted $\hat{m}_{gh}$, we can estimate the response probability with (6)

$$\text{(6)} \qquad \hat{p}_i = \frac{n_g}{n_g + \hat{m}_g} = \frac{\sum_{h=1}^{H} n_{gh}}{\sum_{h=1}^{H} n_{gh} + \sum_{h=1}^{H} \hat{m}_{gh}} \qquad \text{for } i \in s_g$$

We use an iterative algorithm to estimate m_{gh}:

1. Initial values for m_{gh} is chosen, denoted by $m_{gh}^{(0)}$, as for instance

$$m_{gh}^{(0)} = \frac{m_h n_{gh}}{n_h} = \frac{m_h n_{gh}}{\sum_{g=1}^{G} n_{gh}}$$

2. For k=1,2,... the following expression is calculated

$$w_{gh}^{(k)} = \frac{(n_{gh} + m_{gh}^{(k-1)})(\sum_h m_{gh}^{(k-1)})}{\sum_h n_{bh} + \sum_h m_{gh}^{(k-1)}} \qquad \text{and} \qquad m_{gh}^{(k)} = \frac{m_h w_{gh}^{(k)}}{\sum_{g=1}^{G} w_{gh}^{(k)}}$$

3. The algorithm is stopped after 40 iterations (k=40), and we use $\hat{m}_{gh} = m_{gh}^{(k)}$ as an estimate for m_{gh}.

To try to estimate the response probabilities under the informative RHG-model, this algorithm was programmed in SAS and the RHGs and auxiliary groups described above were used in the estimation process. With the described RHGs and auxiliary groups we got no convergence for the algorithm. We tried to define *h* and *g* for different variables to get the algorithm to converge, but with no luck. To carry out this analysis we have nevertheless used the results from the algorithm after 40 iterations. The results from this analysis must therefore be interpreted with an eye to the fact that the response probabilities may not be correct.

The following response probabilities, $\hat{p}_i$, were estimated for the three response alternatives:

Table 6 Estimated response probabilities in per cent

	Better	**Unchanged**	**Worse**
$\hat{p}_i$	85.3	86.0	91.0

From the table with the estimated response probabilities for the three RHGs, we see that the estimates are about the same for those who answer *better* or *unchanged*, while it is somewhat higher for those who answer *worse.*

By using the estimated response probabilities in equation (6), we are able to calculate the non-response-adjusted weight for each unit, which will depend on the response alternative chosen by the unit.

(7) $$\hat{w}_i = a_i \hat{\phi}_i = (\pi_i \hat{p}_i)^{-1}$$

By using the estimated non-response-adjusted weights from (7) in equation (3), we obtain an estimate of the proportion who say that the outlook is better.

(8) $$\hat{\bar{Y}} = (\sum_{i \in s_r} \hat{w}_i y_i)/(\sum_{i \in U} S_i) = \frac{69675.5}{292940} = 0.238$$

As the estimate in (8) shows, this model results in a marginally higher estimate on the proportion who consider the outlook to be better. This follows from the estimated response probabilities. The estimated non-response weight, $\hat{\phi}_i = 1/\hat{p}_i$, will be higher for units responding *better* than for those responding *worse.* This way the proportion *better* increases in relation to the other alternatives – because the non-response is expected to be higher in this group. It is important to take into

consideration that the estimated response probabilities may be wrong because the algorithm, based on our assumptions, did not converge. It is difficult to provide a good reason why the non-response should be lower for units who answer *worse* than units who answer *better*. One possible reason may be that during an economic recession in the industry there is a greater need to complain (through official statistics) than when the economy is recovering. To substantiate this hypothesis one could carry out an analysis of the unit non-response over a period of time, to find out whether the response rate is correlated with the business cycle. This would, however, extend the scope of this analysis.

3.1.1.4 Calibration of direct weighting by use of a ratio estimator

The models tested until now only use information from the sample. By using additional information from the population one may improve the quality of the direct weighted estimate. Several methods can be used to calibrate the simple estimates; post stratification, ratio model, regression model. In this analysis we will use a ratio model.

We define employment, S_i, as an additional variable. With respect to reduction in variance and adjusting for non-response, it is desirable to use an additional variable that is highly correlated with the variable of interest. We cannot be certain of any correlation between the choice of answer and the level of employment, but in lack of another register variable, employment is used. The ratio estimator is then defined by:

$$(9) \qquad w_{i,rat} = w_i(S/\hat{S}) = \frac{(Sw_i)}{\sum_{i \in s_r} w_i S_i} \qquad \text{where} \quad S = \sum_{i \in U} S_i \qquad \text{and} \qquad \hat{S} = \sum_{i \in s_r} w_i S_i$$

Total number of employees, *S*, in the population is known: $S = \sum_{i \in U} S_i = 292940$

and we know S_i for units who have returned the questionnaire.

We use this method of calibration on the three models of weighting for unit non-response that we have analysed:

a) Direct weighting (MCAR non-response)
b) Non-informative RHG-model
c) Informative RHG-model

a) Direct weighting (MCAR non-response)

By using the calibrated non-response-adjusted weight, $w_{i,rat}$, from (9) in equation (3) we get the following calibrated ratio estimate for the proportion who answer *'better'*

$$(10) \qquad \hat{\bar{Y}}_{rat} = (\sum_{i \in s_r} w_{i,rat} y_i)/(\sum_{i \in U} S_i) = (\sum_{i \in s_r} w_i (S/\hat{S}) y_i)/(\sum_{i \in U} S_i) = \frac{72840.1}{292940} = 0.249$$

In the estimation process we find that the ratio $S/\hat{S}$ is 1,065. This way the employment-calibrated estimate is adjusted somewhat up, in relation to the direct weighting, because of too low coverage of employment in the sample on account of the non-response. Without calibration the non-response units will have identical weights $\phi = p_i^{-1} = (n/(n+m))^{-1}$. When we wish that units with a high level of employment should have a greater impact than units with a low level of employment, the ratio model will compensate for this.

b) Non-informative RHG-model

We calculate the calibrated non-response-adjusted weight, $w_{i,rat}$, in the same way as under the assumption of MCAR non-response with equation (9). However, this time one rate is calculated for each RHG. In this example we will calculate for both a) and b)

a) RHG equal to employment stratum

b) RHG equal to industry

By using equation (9) and (3) we obtain the following calibrated ratio estimate of the proportion who answer *'better'*

$$\text{(10a)} \quad \hat{\bar{Y}}_{rat} = (\sum_{i \in s_r} w_{i,rat} y_i)/(\sum_{i \in U} S_i) = (\sum_{i \in s_r} w_i (S/\hat{S}) y_i)/(\sum_{i \in U} S_i) = \frac{72920.2}{292940} = 0.249$$

$$\text{(10b)} \quad \hat{\bar{Y}}_{rat} = (\sum_{i \in s_r} w_{i,rat} y_i)/(\sum_{i \in U} S_i) = (\sum_{i \in s_r} w_i (S/\hat{S}) y_i)/(\sum_{i \in U} S_i) = \frac{73145.6}{292940} = 0.250$$

Also in relation to the non-informative RHG-model, the non-response-adjusted weights are calibrated with the ratio $S/\hat{S}$ = 1.065. This way the employment-calibrated estimates are somewhat higher than is the case of the non-informative RHG-model without calibration.

c) Informative RHG-model

As was the case in the non-informative RHG-model we will estimate the calibrated non-response-adjusted weight, but in this case we use the estimated response probability where the non-response is correlated with the variable of interest. The calibrated non-response-adjusted weight is denoted as $\hat{w}_i = a_i \hat{\phi}_i = (\pi_i \hat{p}_i)^{-1}$. By using this estimate in equation (9) and (3) the following expression is obtained for the calibrated estimate of the informative RHG-model

$$\text{(11)} \quad \hat{\bar{Y}}_{rat} = (\sum_{i \in s_r} \hat{w}_{i,rat} y_i)/(\sum_{i \in U} S_i) = (\sum_{i \in s_r} \hat{w}_i (S/\hat{S}) y_i)/(\sum_{i \in U} S_i) = \frac{74166.3}{292940} = 0.253$$

As was the case with direct weighting with MCAR non-response and the non-informative RHG-model, the employment-calibrated answer is calibrated with the ratio $S/\hat{S}$ = 1.065, which results in an increase in the estimate compared to the estimate without calibration.

3.1.2 Imputation of item non-response

In this section we will investigate methods for imputation of non-response. Instead of weighting, we now try to generate answers for the non-response units and in this way construct a complete dataset for the gross sample (see fig 1).

Two types of imputation:

- Deterministic: The same values are imputed when the imputation process is repeated
- Stochastic: Different values may be imputed when the process is repeated, and consequently the result may be different each time the imputation process is carried out.

3.1.2.1 Nearest-neighbour imputation

This is a deterministic method that estimates response alternatives based on a metric function that uses additional variables to measure the 'distance' between a non-response unit and a donor[7]. Employment is used as an additional variable, S_i. We then get the distance between the non-response unit and the donor by estimating:

(12) $$\delta_{ij} = |S_i - S_j|$$

This way the response alternative is imputed from the unit that generates the smallest possible δ_{ij} between a non-response unit and a potential donor. I.e. the donor with a number of employees closest to the number of employees in the non-response unit. We therefore assume that there is a correlation between which response alternative is selected and number of employees.

From (1) we have the variable of interest $y_i = \beta_i * S_i$

Where $\beta_i = \begin{cases} 1 & \text{If unit } i \text{ has chosen } \textit{'better'} \\ 0 & \text{If unit } i \text{ has chosen a different response} \end{cases}$

and S_i is the number of employees for unit *i*

With imputed values $\beta_i^* = \beta_j$ where $\delta_{ij} = |S_i - S_j|$ is minimized we get the following expression

$$\tilde{\beta}_i = \begin{cases} \beta_i & i \in s_r \\ \beta_i^* & i \in s_m \end{cases}$$

From this expression we get

(13) $$\tilde{y}_i = \tilde{\beta}_i * S_i$$

We now have, including the imputed values, a value for all units in the gross sample. This implies that in the calculation of the estimate, the non-response-adjusted weight is equal to the design weight and the response probability, p_i is equal to 1

(14) $$w_i = a_i \phi_i = (\pi_i p_i)^{-1} = (\pi)^{-1} = a_i$$

To estimate the employment-weighted proportion, $\hat{\bar{Y}}_{imp}$, we use (13), (14) and (3) and get

(15) $$\hat{\bar{Y}}_{imp} = (\sum_{i \in s} a_i \tilde{y}_i)/(\sum_{i \in U} S_i) = \frac{67574.1}{292940} = 0.231$$

From (15) we see that the non-response adjusted estimate based on nearest-neighbour imputation results in a somewhat lower estimate than what we obtained with models based on weighting for non-response. The proportion of imputed values that was assigned the value $\beta_i^* = 1$ was 0.239, but because the response alternatives are weighted with the level of employment the employment-weighted

[7] The unit from which the imputation value is derived from.

estimate is lower. Among those that were imputed with the value $\beta_i^* = 0$ there was an over-representation of larger units (units with a high level of employment). Table 7 illustrates this.

Table 7 Distribution of imputed values

Imputed value β_i^*	Units	Sum employment
1	22	2984
0	70	19400
Sum	92	22384
Proportion	0.239	0.133

From the table we can see that the proportion that was imputed with the value 1 was 0.239, but if we look at the proportion of the employment-weighted imputed response alternative which was assigned the value 1, the figure was only 0.133.

3.1.2.2 Stochastic imputation with a non-informative RHG-model (hot-deck)

In contrast to imputation with 'nearest-neighbour' this method of imputation is stochastic. This means that repeated simulations of the imputation process generate different estimates. In hot-deck imputation the purpose is to group together units that in some way resemble each other.

To group the units we have chosen RHG = Industry (defined by table 5), as we assume that units belonging to the same industry group have a higher probability of having the same business cycle than units belonging to different industries. In this way we will draw a donor from the same industry group => Non-response in the industry: Textiles, wearing apparel and leather, is covered by imputation from a donor in the same industry.

The method of imputation is based on imputing the value β_i^* from a donor drawn randomly within the same RHG. In the same way as for 'nearest-neighbour' we have the expression

$$\tilde{\beta}_i = \begin{cases} \beta_i & i \in s_r \\ \beta_i^* & i \in s_m \end{cases}$$

To estimate the employment-weighted proportion, $\hat{\bar{Y}}_{imp}$, we use (13), (14) and (3) and once again we get

$$(15) \qquad \hat{\bar{Y}}_{imp} = (\sum_{i \in s} a_i \tilde{y}_i)/(\sum_{i \in U} S_i)$$

Because this method of imputation is stochastic, the estimates will vary when the imputation process is repeated. As an estimate for the stochastic estimate we have chosen to run the simulation 20 times and then compute the expected value, given by the average of the stochastic estimates.

$$(16) \qquad \mathrm{E}(\hat{\bar{Y}}_{imp}) \approx \frac{\sum_{i \in N} \hat{\bar{Y}}_{imp,i}}{N} = 0.233 \qquad \mathrm{N}=(1,...,20)$$

From (16) we see that the average of the 20 simulations results in the same estimate as when we used direct weighting with the assumption of MCAR non-response, but the uncertainty in the estimate has increased because of the stochastic process. The results from the 20 simulations are shown in table 8.

As this table shows, the estimates adjusted for non-response by the use of hot-deck imputation varies from 0.220 to 0.257. The reason for this is that we draw at random within each RHG, thus we get different donors every time the simulation is carried out.

Table 8 Results from hot-deck imputation

Simulation	$\hat{\bar{Y}}_{imp}$
1	0.233
2	0.233
3	0.235
4	0.231
5	0.232
6	0.250
7	0.226
8	0.241
9	0.239
10	0.237
11	0.231
12	0.222
13	0.257
14	0.231
15	0.232
16	0.229
17	0.234
18	0.220
19	0,222
20	0,230
Average	0,233
St. dv	0,009

3.1.2.3 Calibration of estimates from imputation models by using a ratio estimator

As was seen in the case of calibration of the direct weighted estimation using a ratio estimator, we can also in the case of imputation carry out a calibration based on additional information from the population. In this case, it is not the non-response-adjusted weights, w_i, that are calibrated, but the design weights, a_i.

From (14) we have $w_i = a_i$ in the case of imputation

By using (9) we can define the calibrated design weight as

(17) $$a_{i,rat} = a_i(S/\widetilde{S}) = \frac{(Sa_i)}{\sum_{i\in s} a_i S_i} \qquad \text{where} \quad S = \sum_{i\in U} S_i \qquad \text{and} \qquad \widetilde{S} = \sum_{i\in s} a_i S_i$$

The total number of employees, *S*, in the population is known: $S = \sum_{i\in U} S_i = 292940$

and we know S_i for all units in the gross sample.

From (17) and (15) we can then define the calibrated employment-weighted estimate based on imputation as

$$(18) \qquad \hat{\bar{Y}}_{imp,rat} = (\sum_{i \in s} a_{i,rat} \tilde{y}_i)/(\sum_{i \in U} S_i) = (\sum_{i \in s} a_i (S/\tilde{S}) \tilde{y}_i)/(\sum_{i \in U} S_i)$$

We use this calibration method on the two methods of imputation described above:

a) Nearest-neighbour
b) Hot-deck

a) Nearest-neighbour

When calibrating the employment-weighted estimate with this method of imputation we get the following result

$$(18) \qquad \hat{\bar{Y}}_{imp,rat} = (\sum_{i \in s} a_i (S/\tilde{S}) \tilde{y}_i)/(\sum_{i \in U} S_i) = 0.241$$

The estimation shows that the ratio $S/\tilde{S}$ is 1.045. In this case too, the employment-weighted estimate is calibrated to be higher, but by a smaller factor than with direct weighting ($S/\hat{S}$ =1.065). This gives the expression

$$(19) \qquad \tilde{S} = \sum_{i \in s} a_i S_i > \sum_{i \in s_r} w_i S_i = \hat{S}$$

This means that the sum of the products of the design weights and employment for all units in the gross sample is higher than the sum of the products of the non-response-adjusted weights and employment for all units in the net sample.

b) Hot-deck

We can also carry out calibration of the employment-weighted estimate based on the ratio estimator when using stochastic imputation with a non-informative RHG-model (hot-deck). As in the example with hot-deck imputation without calibration, we use RHG = Industry (defined by table 5).

By using (18) and (16) we can estimate an average of the calibrated stochastic estimates by running 20 simulations, and then compute the average of the stochastic estimates.

$$(18) \qquad \hat{\bar{Y}}_{imp,rat} = (\sum_{i \in s} a_i (S/\tilde{S}) \tilde{y}_i)/(\sum_{i \in U} S_i)$$

$$(20) \qquad \mathrm{E}(\hat{\bar{Y}}_{imp,rat}) \approx \frac{\sum_{i \in N} \hat{\bar{Y}}_{imp,rat,i}}{N} = 0.244 \qquad \mathrm{N}=(1,...,20)$$

As shown by (20), the average of the 20 simulations results in a somewhat higher estimate than with hot-deck imputation without calibration by the use of a ratio estimator. This is because the ratio, $S/\tilde{S}$, is 1.045. This ratio will be constant (not stochastic) because the ratio does not depend on the response alternatives, $\tilde{\beta}_i$, and the ratio will be the same as in the case of imputation based on 'nearest-neighbour'.

The reason why the ratio is the same in the two cases is that the gross sample, the design weights and the level of employment are the same in the two cases, and independent of $\widetilde{\beta}_i$ (see equation (17)). The difference in the estimates lies in the values which are imputed for the non-response units.

The results from the 20 simulations are shown in table 9. As the table shows the estimates, adjusted for non-response by the use of hot-deck imputation calibrated with a ratio estimator, vary from 0.226 to 0,258. As was the case for hot-deck imputation without calibration we get a stochastic process because the donors are drawn at random within each RHG every time we run the simulation.

Table 9 Results from the ratio calibrated hot-deck imputation

Simulation	$\hat{\bar{Y}}_{imp,rat}$
1	0.226
2	0.246
3	0.250
4	0.239
5	0.258
6	0.233
7	0.241
8	0.240
9	0.249
10	0.247
11	0.243
12	0.230
13	0.239
14	0.258
15	0.236
16	0.246
17	0.256
18	0.247
19	0.255
20	0.240
Average	0.244
St. dv	0.009

3.1.3 The effect of calibration

To investigate the obtained effect of calibration by the use of the ratio estimator, we can analyse the variance of the employment-weighted estimate with and without calibration. To obtain reduction in the variance of the estimate – calibrated with the use of the ratio estimator – the additional variable, employment, should be correlated with the variable of interest. This is not an unreasonable assumption as the variable of interest y_i is the product of the employment level of the unit and the answer to the question (1 or 0). To investigate if the ratio estimator we have used in the calibration produces any reduction in variance, we measure the effect of additional information conditioned on the adjustment of non-response, i.e. the non-response weights. From Zhang (2003) we have a definition of the estimate of variance for the direct weighted estimator $\hat{\bar{Y}}$, where we assume constant variance:

$$(21) \qquad v_1 = \frac{1}{n}(1 + c_w^2)s_y^2$$

where c_w^2 is the coefficient of variance to w_i over s_r, and var(y_i), denoted s_y^2, may be written as

$$(22) \qquad s_y^2 = \frac{1}{n-1}\sum_{i \in s_r}(y_i - \bar{y})^2$$

where y_i is the employment-weighted answer defined as in equation (1), and the average $\bar{y}$ is

$$(23) \qquad \bar{y} = \frac{1}{n}\sum_{i \in s_r} y_i$$

This will hold regardless of the non-response model being informative or not.

A simple variance estimate for the calibrated estimator, under similar assumptions, has the following general expression

$$(24) \qquad v_2 = \frac{1}{n}(1 + c_w^{*2})s_e^2$$

where c_w^{*2} is the coefficient of variance to the calibrated weights, and s_e^2 is the variance of the calibration residuals. The definition of the calibration residuals given ratio estimation is

$$(25) \qquad e_i = y_i - x_i\beta = y_i - x_i\frac{\hat{Y}}{\hat{X}} = y_i - x_i\frac{\sum_{i \in s_r} w_i y_i}{\sum_{i \in s_r} w_i x_i}$$

In equation (25) the additional variable employment, which is used in the ratio, is denoted x_i. The variance of the calibration residuals can then be calculated by the following formula

$$(26) \qquad s_e^2 = \frac{1}{n-1}\sum_{i \in s_r}\left(y_i - x_i\frac{\sum_{i \in s_r} w_i y_i}{\sum_{i \in s_r} w_i x_i}\right)^2$$

With these coherences we can measure the effect of additional information by the ratio:

$$(27) \qquad \eta = \frac{v_2}{v_1} = \frac{1 + c_w^{*2}}{1 + c_w^2}.\frac{s_e^2}{s_y^2}$$

Assuming that non-response is MCAR, and that the coefficient of variance for the calibrated weights is approximately equal to the coefficient of variance for the non-response-adjusted weights without calibration, $c_w^* \approx c_w$, we can reduce the ratio to $\eta \approx s_e^2 / s_y^2$. We have calculated this ratio for the model with direct weighting and MCAR non-response, with and without calibration. The result from this calculation is expressed in (28).

$$(28) \qquad \eta \approx \frac{s_e^2}{s_y^2} = \frac{33320}{44982} = 0.74$$

In other words, we reduce the variance in the employment-weighted estimate by 26 per cent by using calibration with the ratio estimator.

4. Summary

In this paper we have investigated the non-response in the Norwegian Business Tendency Survey for manufacturing, mining and quarrying, and in particular question 18; *General judgement of the outlook for the establishment in the next quarter.* Chapter 3 presented a general description of possible reasons for unit and item non-response. Aggregated response rates for the four employment strata are also calculated (see table 3).

In chapter 3.1, Adjustment for non-response, we have adjusted the employment-weighted estimate, for the proportion who believe that the general outlook is better, by using different models for weighting for non-response and two different methods of imputation. In addition to this we have used a ratio estimator to calibrate these estimates. In chapter 3.1.3 we have investigated the effect of calibration with the use of the ratio estimator, and in particular if it generates any reduction in variance.

The estimate using direct weighting with the assumption of MCAR non-response, calibrated with a ratio estimator, is approximately the same as the one generated by the current quarterly production of the statistics. In the current production of the statistics we assume MCAR non-response and in the ratio estimator only the net sample is included. One distinction is that in the quarterly production process the item non-response is calculated as a separate response alternative (denoted proportion 'Non-response'). Another distinction is that we calibrate with the ratio for each employment strata in each industry.

If we look at the estimates from the non-informative RHG-model, the results are approximately the same as under the assumption of MCAR non-response. This applies both for the estimates with and without calibration with the use of the ratio estimator. This indicates that the definition of the response homogeneity groups (employment strata and grouping by industry) does not produce groups with different patterns of non-response between the groups, and therefore no adjustment of the estimates is recorded. There may be other ways of grouping the RHGs, in such a way that the response probability is different between the different groups, and as equal as possible within the group. However, we have not been able to find such a classification.

The results from the informative RHG-model are somewhat higher than for the rest of the estimates, which indicate that the level of non-response is higher among the units that expect a better development in the next quarter. Because the algorithm did not converge, this has an effect on the estimates and makes it difficult to draw any conclusions.

In chapter 3.1.2, Imputation of item non-response, we use two different methods of imputation; one deterministic (nearest neighbour) and one stochastic (hot-deck). The result using 'nearest neighbour' imputation shows that this estimate is somewhat lower than the results from the models of weighting for non-response. This indicates that there was an overrepresentation of larger units who got the imputed value 0 (response alternative *'unchanged'* or *'worse'*). This way the employment-weighted estimate for the proportion who believe that the general outlook is *'better',* is somewhat reduced. When it comes to the stochastic imputation with a non-informative RHG-model (hot-deck), we see that the average estimate, based on 20 simulations, is the same as under the assumption of MCAR non-response. When RHG equal to industry group proved to have little impact when used in the non-response model, this method of imputation will produce an estimate corresponding to imputation using random donors within the whole net sample (no RHG-index). Given that these results are equal, the non-response model with MCAR non-response would be preferred, because this estimation process will generate a lower level of variance than is the case of the stochastic imputation procedure.

Further we note that in the case of calibration with ratio estimation, we get a lower estimate with the methods of imputation than is the case for non-response models. Table 10 sums up the different estimates we have calculated.

Table 10 Results from adjusting for non-response with the use of non-response models and methods of imputation

		Non-response model				Method of imputation	
		MCAR non-response	Non-informative RHG		Informative RHG	Nearest neighbour	Hot-deck (RHG=Industry)
			Empl. strata	Industry group			
Calibration with ratio estimator	No	0.233	0.234	0.234	0.238	0.231	0.233
	Yes	0.249	0.249	0.250	0.253	0.241	0.244

From the calculations that have been carried out it is difficult to conclude that there is variation in the distribution of non-response in the Norwegian Business Tendency Survey, but we cannot rule out the possibility that mechanisms of non-response causes systematic variation.

Calibration of the estimates with the ratio estimator results in a higher proportion who consider the general outlook to be better. This holds for all the investigated cases. The calculation of the effect of calibration with the use of ratio estimation shows that this procedure generates estimates with a lower level of variance than estimates without calibration, and that it is reasonable to calibrate the estimates in this way.

We have made a number of simplifications in this analysis. For instance, we only look at one response alternative and one question. A number of the questions in the Business Tendency Survey are correlated with each other, and this has an impact on which methods should be used to adjust for non-response. The advantage of the estimation procedure used in the quarterly production of the statistics, with the assumption of MCAR non-response, is that it provides a simple and straightforward method for generating results for all questions as a whole.

If further analysis should be carried out in relation to the non-response in the Business Tendency Survey, it would be interesting to take a closer look at different methods of imputation. For multi-purpose estimation, a method of imputation is easier to be integrated into the existing production process than the weighting adjustments.

References

Zhang (2003): SM05 - Introduction to adjusting for non-response, unpublished paper for the course SM05 at Statistics Norway August 2003, Li-Chun Zhang, (in Norwegian only)

Recent publications in the series Documents

2001/12 B. Hoem: Environmental Pressure Information System (EPIS) for the household sector in Norway

2001/13 H. Brunborg, I. Bowler, A.Y. Choudhury and M. Nasreen: Appraisal of the Birth and Death Registration Project in Bangladesh

2001/14 K. Rypdal: CO_2 Emission Estimates for Norway. Methodological Difficulties

2001/15 E. Røed Larsen: Bridging the Gap between Micro and Macro: Interdependence, Contagious Beliefs and Consumer Confidence

2001/16 L. Rogstad: GIS-projects in Statistics Norway 2000/2001

2002/1 B. Hoem, K. Erlandsen og T. Smith: Comparisions between two Calculation Methods: LCA using EPIS-data and Input-Output Analysis using Norway' s NAMEA-Air Data

2002/2 R. Bjørnstad: The Major Debates in Macroeconomic Thought - a Historical Outline

2002/3 J. L. Hass and T. Smith: Methodology Work for Environmental Protection Investment and Current Expenditures in the Manufacturing Industry. Final Report to Eurostat.

2002/4 R. Bjørnstad, Å. Cappelen, I. Holm and T. Skjerpen: Past and Future Changes in the Structure of Wages and Skills

2002/5 P. Boug, Å. Cappelen and A. Rygh Swensen: Expectations and Regime Robustness in Price Formation: Evidence from VAR Models and Recursive Methods

2002/6 B.J. Eriksson, A.B. Dahle, R. Haugan, L. E. Legernes, J. Myklebust and E. Skauen: Price Indices for Capital Goods. Part 2 - A Status Report

2002/7 R. Kjeldstad and M. Rønsen: Welfare, Rules, Business Cycles and the Employment of Single Parents

2002/8 B.K. Wold, I.T. Olsen and S. Opdahl: Basic Social Policy Data. Basic Data to Monitor Status & Intended Policy Effects with Focus on Social Sectors incorporating Millennium Development Goals and Indicators

2002/9 T.A. Bye: Climate Change and Energy Consequenses.

2002/10 B. Halvorsen: Philosophical Issues Concerning Applied Cost-Benefit Analysis

2002/11 E. Røed Larsen: An Introductory Guide to the Economics of Sustainable Tourism

2002/12 B. Halvorsen and R. Nesbakken: Distributional Effects of Household Electricity Taxation

2002/13 H. Hungnes: Private Investments in Norway and the User Cost of Capital

2002/14 H. Hungnes: Causality of Macroeconomics: Identifying Causal Relationships from Policy Instruments to Target Variables

2002/15 J.L. Hass, K.Ø. Sørensen and K. Erlandsen: Norwegian Economic and Environment Accounts (NOREEA) Project Report -2001

2002/16 E.H. Nymoen: Influence of Migrants on Regional Varations of Cerebrovascular Disease Mortality in Norway. 1991-1994

2002/17 H.V. Sæbø, R. Glørsen and D. Sve: Electronic Data Collection in Statistics Norway

2002/18 T. Lappegård: Education attainment and fertility pattern among Norwegian women.

2003/1 A. Andersen, T.M. Normann og E. Ugreninov: EU - SILC. Pilot Survey. Quality Report from Staistics Norway.

2003/2 O. Ljones: Implementation of a Certificate in Official Statistics - A tool for Human Resource Management in a National Statistical Institute

2003/3 J. Aasness, E. Biørn and t. Skjerpen: Supplement to <<Distribution of Preferences and Measurement Errors in a Disaggregated Expenditure System>>

2003/4 H. Brunborg, S. Gåsemyr, G. Rygh and J.K. Tønder: Development of Registers of People, Companies and Properties in Uganda Report from a Norwegian Mission

2003/5 J. Ramm, E.Wedde and H. Bævre: World health survey. Survey report.

2003/6 B. Møller and L. Belsby: Use of HBS-data for estimating Household Final Consuption Final paper from the project. Paper building on the work done in the Eurostat Task Force 2002

2003/7 B.A. Holth, T. Risberg, E. Wedde og H. Degerdal: Continuing Vocational Training Survey (CVTS2). Quality Report for Norway.

2003/8 P.M. Bergh and A.S. Abrahamsen: Energy consumption in the services sector. 2000

2003/9 K-G. Lindquist and T. Skjerpen: Exploring the Change in Skill Structure of Labour Demand in Norwegian Manufacturing

2004/1 S. Longva: Indicators for Democratic Debate - Informing the Public at General Elections.