A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 235, 2010

*Barbara Kowalczyk**

ON SELECTED ISSUES OF NON-RESPONSE IN THE CASE OF BUSINESS TENDENCY SURVEYS

Abstract. The paper provides empirical analysis of different methods of handling non-response adopted to the very special case of business tendency surveys. We propose reweighted methods of estimation of balance statistics and methods of data imputation. Empirical analysis is based on Business Tendency Surveys (BTS) conducted by the Research Institute for Economic Development (RIED) of the Warsaw School of Economics. Business surveys are conducted in order to obtain information about economic situation and are of significant importance in economic research.

Key words: business tendency surveys, non-response, weighting methods, imputation, balance statistics.

I. BUSINESS TENDENCY SURVEYS AND DESCRIPTION OF RIED DATA

Business surveys are conducted in order to obtain information about economic situation. The main advantages of business research are rapidity of collecting information and that it gives an up-to-the-minute account of business situation. Tendency surveys, that is qualitative business surveys, although do not provide so many possibilities of formal analysis as compared to quantitative business surveys, are often regarded as more reliable. Respondents are more likely to correctly identify direction of changes than to provide a precise point assessment or forecast. Therefore in many analyses, economists purposefully use qualitative measurements obtained from business tendency surveys.

The most commonly employed aggregated measures of survey respondents in tendency surveys are balance statistics (balances). Balances assigned to each survey question are defined as differences between shares of "optimistic" and "pessimistic" responses. If we denote shares of respondents reporting improvement noted in period $t+k$ in comparison to period t by ${}_tI_{t+k}$ and shares

* Ph.D., Institute of Econometrics, Warsaw School of Economics, e-mail: Barbara.Kowalczyk@sgh.waw.pl

of respondents reporting deterioration noted in period $t + k$ in comparison to period t by ${}_tD_{t+k}$, then balance statistic is defined as:

$$ {}_tB_{t+k} = {}_tI_{t+k} - {}_tD_{t+k} . \tag{1}$$

In construction of more advanced balance statistics respondents are often weighted in order to account for different significance of respondents. In case of industrial enterprises the most standard weights are size of employment or value of turnover.

Business Tendency Surveys conducted by the Research Institute of Economic Development (RIED) at the Warsaw School of Economics launched in 1986. In monthly surveys of industrial enterprises respondents assess changes in eight selected areas of economic activity: production, orders, export orders, stock, prices, employment, financial standing, general situation of the economy. They may choose one of the following three options: improvement, no changes, deterioration. Only as for export orders, a variant "do not apply" is added. Both current situation and forecasts are assessed according to this pattern. For all questions balance statistics are calculated. As it is seen from (1) option "no changes" does not influences balances. In case of RIED survey balances are weighted with employment numbers and may vary from –100 to +100. More precisely, five employment levels are distinguished: up to 50 persons, 51 to 250 persons, 251 to 500 persons, 501 to 2000 persons, over 2000 persons. The answers are weighted by 1, 2, 3, 4 and 5 respectively, depending on to which employment class the given industry enterprise belongs (see RIED 2008).

.

II. UNIT NON-RESPONSE AND THE STRUCTURE OF THE SAMPLE

The main advantages of business research, rapidity of collecting information and the fact that it gives an up-to-the-minute account of business tendencies, are connected with the main disadvantage of the survey, that is a large unit non-response rate, i.e. the rate of manufacturing firms which have not returned their questionnaires. In business tendency surveys there is no time for repeated mail, reserve sample, subsampling of non-respondents or any other methods commonly used in spread out in time surveys to avoid non-response.

Key issues concerned with the unit non-response problem are: evaluation of representativeness of the sample and changes of the structure of the sample in time. Unfortunately employment levels available from The Central Statistical Office of Poland: to 9 persons, 10 – 49 persons, 50-249 persons, 250-999 persons, 1000 and more persons do not agree with employment levels available in RIED data. Combining employment strata available from The Central

Statistical Office data (GUS, 2007) and employment strata available in RIED sample (RIED, 2008) we compute comparable structures of the population and the sample regarding employment levels. Results are presented in table 1.

Table 1. Structure of the population and of RIED sample

Employment level	Population 2007	Sample 01.2008	Sample 02.2008	Sample 03.2008	Sample 04.2008
To 50 persons	97,47%	38,46%	41,83%	28,16%	36,00%
51-250 persons	2,10%	36,00%	33,99%	35,08%	36,25%
Over 250 persons	0,43%	25,54%	24,18%	36,76%	27,75%

Source: Own calculations on the basis of GUS (2007) and RIED data base.

One issue has to be emphasized while interpreting table 1. Although manufacturing firms with employment over 250 persons constitute only 0,43% of all firms in the population, they are of utmost importance as far as contribution to the GDP is concerned. Higher significance of a large single industry as compared to the smaller one has it's reflection in weights making allowances for employment levels used in constructing balance statistics (see section 1).

Now we will conduct an analysis of estimating the population parameter defined as the balance statistics weighted with employment numbers computed for all manufacturing firms in Poland. To focus attention we will analyze only two questions in the survey: production and general economy situation – both state and forecasts. We will use data from the surveys carried out in 4 succeeding months from January to April 2008. Two main issues connected with unit non-response are the following: sample does not reflect the structure of the population and the structure of the sample vary from one month to another. Reweighting is one of the method used to adjust for non-response. In constructing reweighted balance statistics we use population weights:

$$W_1 = \frac{N_1}{N} = \frac{365526}{375015} = 0{,}9747,$$

$$W_2 = \frac{N_2}{N} = \frac{7864}{375015} = 0{,}021,$$

$$W_3 = \frac{N_3}{N} = \frac{1625}{375015} = 0{,}0043.$$

Results of reweighting as compared to original RIED method is given in table 2.

Table 2. RIED and reweighted balances

Period	Production state		Production forecast		General economy situation state		General economy situation forecast	
	RIED balance	Reweighted balance	RIED balance	Reweighted balance	RIED balance	Reweighted balance	RIED balance	Reweighted balance
2008.01	-7,6	-12,7	16,2	11,5	4,2	6,1	3,2	9,8
2008.02	-3,6	-3,7	22,0	19,7	4,5	8,6	3,6	13,5
2008.03	23,0	3,3	35,3	26,9	13,8	8,7	15,4	11,4
2008.04	2,7	3,2	21,6	23,6	-0,9	1,4	1,0	-2,4

Source: Own calculations on the basis of RIED data.

One has to remember that the main purpose of tendency surveys is to observe tendencies. All tendencies in analyzed period agree in the two methods of computing balances except for a change between February and March as far as forecast for general economy situation is concerned. The lack of agreement of tendencies is well seen in figure 1.

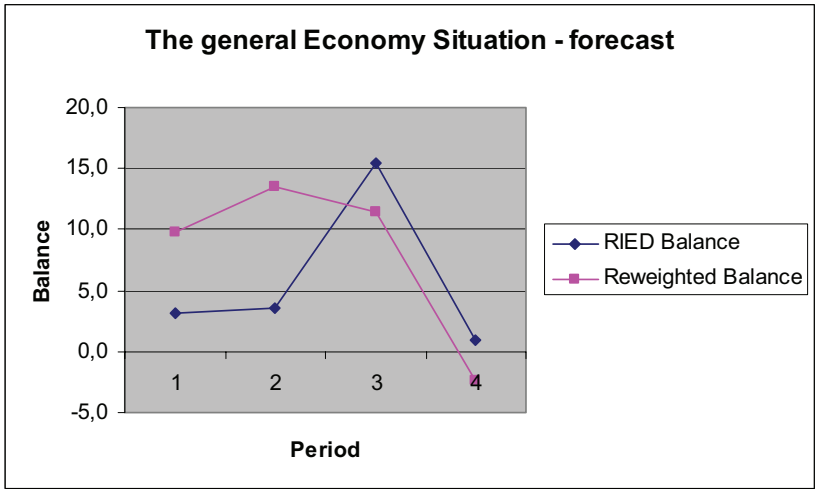

Figure 1. RIED and reweighted balance for general economic situation - forecast

The reason for the situation can be easily explained. In March as compared to February two changes has occurred: the structure of the sample has changed significantly (see table 1, perceptible increase of large manufacturing firms) and higher optimism among large manufacturing firms has been observed as far as forecast for general economy situation has been concerned.

III. ITEM NON-RESPONSE AND IMPUTATION METHODS

Item non-response is typically handled by imputation methods. Imputation implies that a missing value of the study variable for a sample element in the data matrix is substituted by an imputed (predicted) value. Properly chosen method of imputation is advisable in many situations. Most common are listed below.

- Even if particular variables separately are characterized by small non-response rates, on aggregated level large numbers of elements may by eliminated because of individual missing items.
- Most analysis tools are suitable only for complete data (rectangular set). The existence of missing values means that the standard procedures cannot be automatically used. Data reduction and data imputation are the two approaches to the problem.It is rare that the missing values occur completely at random, as they might in some experimental contexts; in surveys, it is often reasonable to suspect that non-respondents systematically differ from respondents, and thus it is desirable to study the effects of various differences between respondents and non-respondents.

Because of variety of imputation methods, decision about choosing one of them should be taken after settlement of some aspects of the survey, such as: non-response rates, mechanism that lead to missing data, aim of the analysis etc. Imputation methods are presented in detail in Longford (2005) and Rubin (2004). Some hitherto experiences and analyses for business surveys (e.g. Norwegian BTSs, Japanese TANKAN) are summarized in Kowalczyk (2008). Imputation methods applied suitable for other business surveys BSs (see Wang 2004, Donzè 2000) cannot be automatically transposed into RIED BTSs, mainly because auxiliary variables available for RIED BTSs are of qualitative nature only, like the level of employment. The exact number of employees is not available for RIED data.

Now we introduce the imputation methods proposed by the Author. Then we apply it to RIED BTS and analyze the results. To focus attention and simplify the presentation of results we analyze variable “general economy situation” - state. The item non-response problem is the most evident in the question about general economy situation, as seen in table 3. Analysis is conducted for April 2008.

Table 3. Item non-response rates in April 2008

Question	State - item non-response	Forecast - item non-response
Level of production	1,0%	2,3%
Level of orders	0,2%	1,3%
Level of export orders	5,2%	6,4%
Stocks of finished goods	2,9%	4,3%
Prices of goods produced	1,5%	2,5%
Level of employment	0,4%	1,3%
Financial standing	0,2%	1,1%
General economy situation	6,9%	8,4%

Source: own calculations on the basis of RIED data.

Model 1. Hot deck class imputation method – non-informative Response Homogeneity Group RHG model. Our aim is to group together units that are similar and resemble each other. To group the units we have chosen both nace and employment level. **So we assume that the highest probability of having the same business cycle have manufacturing firms representing the same nace and similar employment.** There are 23 naces and 5 employment levels thus we have 115 response homogeneity groups RHGs (classes). For an element in a given class, an imputed value is obtained from a randomly chosen donor in the same class. The method is stochastic, data imputed will vary when the imputation process is repeated, 10 independent repetitions within this model are conducted**.**

Model 2. Nearest neighbor imputation method. For a subject whose record is incomplete we define the nearest neighbor as a subject from the units with completed records with the maximum number of questions identically answered. As usually there are several nearest neighbor subjects, a donor is chosen among them randomly, so in our case the method is also stochastic and can be treated as another class hot deck method; 10 independent repetitions within this model are conducted.

Assumption under model 1 has very strong economic background. Model 2 is informed by other variables' values. Under both models multiple imputation has been conducted, as described in Rubin (2004). For each set of imputations a complete data set is created. Table 4 presents the estimates associated with each of obtained data sets.

Table 4 Results of multiply imputed data set for RIED sample

Repetition	Balance under model 1	Balance under model 2
1	–1,26	–0,52
2	–0,31	–0,84
3	–1,26	–0,63
4	–1,36	–0,31
5	–0,10	–1,15
6	–1,47	–0,63
7	–0,21	–0,94
8	–1,15	–0,63
9	–1,05	–1,15
10	–1,15	–0,94

Source: own calculations on the basis of RIED data.

As we see from table 4 estimates by the use of hot-deck imputation varies for the balance statistics from –1,47 to –0,1. In the case of nearest neighbor imputation the range of the results varies from –1,15 to –0,31.

Next, we will combine the results received under the same model to obtain one inference under the model, and, to this effect we will derive averages of the results (estimates). The results, along with the between-imputation standard deviation of the estimate are given in table 5 and are compared with the results obtained on the original RIED sample (without filling in empty items). From table 5 we see that the average of the 10 imputation result in very similar estimates to those obtained on the original RIED sample. Nevertheless it should be emphasized here that estimation is not the main reason for which imputation is usually done. The main applications of imputations are listed in the beginning of the section. But still comparison of the estimates is advisable.

Table 5. Combined estimates of balance statistic from table 3 along with the between imputation standard deviation of the estimate and balance based on original RIED data

	Original sample	Model 1	Model 2
Estimate	-0,88	-0,93	-0,77
Between imputation standard deviation	X	0,52	0,28

Source: Own calculations.

After conducting similar computations, estimates of balance statistics for the period from January 2008 to April 2008 have been obtained. Results are summarized in table 6.

Table 6. Balance statistics

Period	Original sample	Model 1	Model 2
01.2008	4,17	4,16	4,15
02.2008	4,52	4,61	4,72
03.2008	13,78	13,99	13,71
04.2008	-0,88	-0,93	-0,77

Source: Own calculations.

Imputation methods presented above are an example of the classical approach to item non-response problems. Some particular problems connected both with item non-response and business surveys can be successfully solved by mathematical procedures. The example is an analysis of rationality of expectations on the basis of business tendency surveys with item non-response, see Kowalczyk, Tomczyk (2008).

IV. CONCLUSIONS

Non-response is present in almost all surveys but the extent and the effect of the non-response can vary greatly from one type of survey to another. Special attention should be dedicated to missing data before inferences and conclusions are drawn. Disregarding of the missing data problems may lead to obtaining results which are biased and incredible. There is no one simple solution of all the problems connected with non-response. The choice of the proper method depends greatly on the mechanisms that lead to missing data, non-response rates, aim of the analysis and auxiliary information available. An attempt to employ techniques of different weighting procedures and imputation methods to results of business tendency surveys can be considered a step towards obtaining broader and consistently more reliable conclusions in many economic and statistical analysis.

REFERENCES

Donzè L. (2000) *Imputation to correct the item nonresponse for the KOF/ETHZ-Innovation Survey 1999,* 25th CIRET Conference Proceedengs, Paris.

GUS [2007], *Zmiany strukturalne grup podmiotów gospodarki narodowej w 2007 r.*, Warszawa 2008.

Kowalczyk B., Tomczyk E. (2008), Rationality of Expectations of Industrial Enterprises – Analysis Based on Business Tendency Surveys with Item Non-response, *Bank i Kredyt*, Vol.8/2008, pp.3-11.
Kowalczyk B. (2008), Application of Imputation Methods in Business Tendency Surveys, Internal report No 03/S/0074/08, Warsaw School of Economics.
Longford N. (2005), *Missing Data and Small-Area Estimation*, Springer Science, Business Media, New York.
RIED (2008), *Business Survey*, May, Research Institute for Economic Development, Warsaw School of Economics, Warsaw.
Rubin D. (2004), *Multiple Imputation for Nonresponse in Surveys,* Wiley, New York.
Wang H. (2004), Non-response in the Norwegian Business Tendency Survey, Department of Economic Statistic Document, No 2004/3, Statistics Norway,
http://www.ssb.no/english/subjects/08/05/10/doc_200403_en/doc_200403_en.pdf

Barbara Kowalczyk

WYBRANE ZAGADNIENIA ZWIĄZANE Z BRAKIEM ODPOWIEDZI W PRZYPADKU BADAŃ KONIUNKTURY

W pracy przedstawione są różne metody podejścia do problemów braków odpowiedzi w przypadku badań koniunktury. Zaproponowane są ważone metody szacowania statystyki bilansowej dla badanych cech oraz metody imputacji danych. Analiza empiryczna oparta jest na Badaniu Koniunktury w Przemyśle prowadzonym przez Instytut Rozwoju Gospodarczego SGH. Badania koniunktury w przemyśle, prowadzone w celu uzyskania informacji o stanie gospodarki, mają kluczowe znaczenie dla praktyki ekonomicznej i stanowią podstawę wielu empirycznych analiz ekonomicznych.